REVIEWS OF MODERN PHYSICS

Volume 29
1957

Published for the

American Physical Society

by the

American Institute of Physics
Incorporated

REVIEWS OF MODERN PHYSICS

Reviews of Modern Physics is published quarterly to provide those interested in developments in physics with comprehensive and timely discussions of the problems which are of especial interest and importance.

It is hoped that the studies in this journal will give to the specialist the benefit of the viewpoint of other specialists; to the student, a background of critical knowledge and an adequate bibliography of source material; to those whose time is absorbed in the teaching of physics, a stimulating account of the forward movement of their science.

Published by the American Institute of Physics for the

AMERICAN PHYSICAL SOCIETY

Manuscripts for publication should be submitted to J. W. Buchta, University of Minnesota, Minneapolis 14, Minnesota.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publication Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price

	U. S. and Canada	Elsewhere
To members of the American Physical Society	$3.00	$3.40
To all others	4.00	4.40

Back Numbers

Yearly rate when complete year is available: **$8.00**
Single copies: **$2.25** each.
Combined issues: **$4.50**

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Changes of address should be addressed to the American Institute of Physics.

Reviews of Modern Physics is published quarterly in January, April, July, and October at Prince and Lemon Streets, Lancaster, Pennsylvania.

Entered as second-class matter December 27, 1945, at the post office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized December 27, 1945.

CONTENTS

JANUARY, 1957

APRIL, 1957

JULY, 1957

Papers from the Conference on the Role of Gravitation in Physics Held at the University of North Carolina, Chapel Hill, North Carolina, January 18–23, 1957

OCTOBER, 1957

REVIEWS OF MODERN PHYSICS

VOLUME 29, NUMBER 1 JANUARY, 1957

Announcement

The Council of the American Physical Society announces that, with the April issue, the *Reviews of Modern Physics* will enter a new phase of its history. The *Reviews*, founded in 1929 by a group of prominent physicists headed by J. T. Tate, will pass from the able and forward-looking leadership of J. W. Buchta into the equally capable hands of E. U. Condon. Dr. Buchta edited the *Reviews* for about twenty years, has brought it to its present position of prominence, and has earned the sincere gratitude of the community of physicists. He retires from the editorship because of the weight of his other commitments to his school and to the Government.

The new editor, Professor Condon, of Washington University, is known to the readers of the *Reviews* as the author of three articles on the "Quantum Mechanics of Collision Processes" (with P. M. Morse, 1931), "Theories of Optical Rotatory Power" (1937), and "Principles of Microwave Radio" (1942). He is known even better for his original research in a great variety of fields and to many of us as an inspiring teacher and colleague. His monumental book, *The Theory of Atomic Spectra*, written in collaboration with G. Shortley in 1935, continues to be the most comprehensive account of our understanding of atomic spectra—the fruit of several decades' intensive study by a great number of physicists of all countries.

We hope that Dr. Condon will be given the same wholehearted support which made the work of his predecessors so rewarding. Manuscripts for the April, 1957, and later issues of the *Review of Modern Physics* should be sent to him, care of the Department of Physics, Washington University, St. Louis, Missouri.

Astronomical Time

G. M. CLEMENCE

U. S. Naval Observatory, Washington, D. C.

INTRODUCTION

IT is well known by this time that the operational approach to science has been exceedingly fruitful, the special theory of relativity being perhaps the outstanding example of a contribution to knowledge that resulted from consideration of the actual operations performed by scientists. It is my purpose here to try to describe the determination of time in terms of the operations involved, and to do it in language intelligible to physicists. The attempt has not been made before, so far as I am aware. The novelty of the approach makes it somewhat difficult to select the topics to be discussed, and to decide what degree of refinement is desirable. Readers who find the going difficult may be helped by reference to two earlier articles on the same subject (Clemence, 1952, 1956),* in which a somewhat different point of view is taken.

I hope that the present attempt may conduce to a more general understanding of the special theory of relativity, which appears to hold some mysteries still for about half the persons who write about it. That this statement is true is demonstrated by the frequent appearance in scientific journals of articles in pairs, where the two authors take opposite sides. In many of these articles the issue involves something that is supposed to be done with a clock, and it is evident that at least one of the authors has no clear idea of what a clock is.

The operational approach is, in practice if not in definition, to some extent a matter of degree. We may describe operations that actually have been performed, or operations that could easily be performed but have not been, or operations that could be performed only with very great effort, or operations that cannot be performed. Among the last we may further distinguish operations that may sometime become possible and those that will not. At one end of the scale we are always sure of our ground but find it difficult to extend our horizon; at the other the horizon becomes boundless but there is nowhere to stand. It is at the latter end that many of the silly questions of students, as well as most of the famous paradoxes of science, originate. The examples of the earth that is supposed to stop rotating and of Maxwell's demon come readily to mind. In my opinion it is better, when introducing a novel subject, to stick pretty closely to the facts; and that is what I shall do here. In the interest of clarity I shall, however, take the liberty of simplifying the actual operations greatly. For example, no mention will be made of the moon as a clock, although it is an indispensable means of subdividing and refining the measurements of the revolution of the earth around the sun. The various noninertial frames of reference that are in use merely to serve the convenience of observers, will not be described. The rotation of the earth, although it served as the fundamental standard of time until a few years ago, and is still extensively used as an adjunct to the new fundamental standard, will be passed over with slight mention. My aim is to describe only the basic concepts in the simplest possible operational terms.

TIME AND TIME INTERVALS

The measurement of time is, basically, the counting of the repetitions of a recurring phenomenon. No measuring rod is required. The measurement of lengths always involves the comparison of two lengths: the length to be measured and the unit of length, whereas the measurement of time requires no such comparison, but merely the counting of the units themselves as they recur. The length of any object is obtained by counting the number of times that the unit of length is contained in it, whereas the time of any event is given by stating the actual count of units of time that has been arrived at when the event is observed to occur. Thus, a determination of time always involves a judgment of simultaneity.

While in principle the measurement of time involves only counting, in practice measurements of lengths are introduced in connection with determinations of time for two different purposes. One is as an aid in the subdivision of the unit of time into parts too small to be immediately perceptible to the senses. The other is as an aid in fixing an epoch with the utmost precision. In neither case is the fundamental principle affected. Probably measurements of lengths are introduced for convenience rather than necessity; microscopes have reached a higher state of development than magnifiers of time intervals. The latter are, however, used to some extent. One method of timing solar eclipses is by taking motion pictures at high speed simultaneously of the sun and a clock, the judgment of simultaneity being made by identifying the first frame on which no portion of the sun appears.

The origin of time, that is, the instant when the count of units is supposed to have begun, is called the fundamental epoch. It is completely at our choice, and is determined by custom and consent. It is common practice to count years from the beginning of the Christian era, months from the (arbitrary) beginning of the year, days from the beginning of the month, and

* See the list of references at the end of the paper.

hours, minutes, and seconds similarly. The vagaries of the calendar are due chiefly to the incommensurability between the day and the year.

Just as the meter sticks in common use are more or less accurate copies of the standard meter, so our ordinary clocks may be regarded as more or less accurate copies of the standard of time.

If we define an epoch as any specified instant of time, then the duration between two distinct epochs is an interval of time. Whenever the time at which an event occurred is stated, an interval of time is implied: the interval between the fundamental epoch and the epoch of the event. In an important sense a determination of time is a determination of the interval of time since the fundamental epoch. So it may be said that we do not measure time but only intervals of time. The important distinction is that when the time of an event is stated, the other end of the interval (the fundamental epoch) is known, whereas when an interval of time is stated its place in the scale of time remains indeterminate. For example, if I say that I visited a friend for half an hour, less information is conveyed than if I say that I visited him from 8:00 until 8:30 P.M. If, in addition, I state the day, month, and year, nothing further can be said about the time of the visit.

VERIFICATION OF EXPERIMENTS

In the making of many experiments, particularly those involving the determination of the rate of change of physical magnitudes with the time, the physicist imagines that he has no need of a clock; a stop watch seems to serve him equally well or better. What he requires is a known interval of time, which need not be fitted in its place within the entire scale of time. The reason he has no interest in the whole scale of time appears to be that he is accustomed to take his intervals of time on faith from the astronomer; experience has shown that the intervals of time so obtained are much more accurate than any measurement the physicist usually makes. Very likely it is for this reason that the time is usually regarded as the independent variable whenever relations are established between it and other variables. But there is no basic reason why determinations of time intervals should be more accurate than other measurements, and in fact they may not always be so. This fact has an important bearing on the verification of experiments. Evidently an experiment in which a time interval plays an important part cannot be said to be completely verified unless the time interval itself is verified.

It is well known that the difference between a scientific experiment and a nonscientific one is that the scientific experiment may be verified, the verification nearly always being made at a different place and time. In verifying the length of a solid object it is usually considered sufficient to measure it a second time, of course with suitable controls of temperature, errors of measurement, and other sorts of variables. An important question to ask is how an interval of time may be verified. Once it is past, the same interval cannot again be measured. How then can we be sure that a second measured with a stop watch today bears any known relation with a second measured fifty years ago? A partial answer to the question is that most of our experiments today give results consistent with those of fifty years ago. But a more precise answer is possible. Our confidence in our determinations of time, and of intervals of time, is determined solely by our ability to calculate the time of past and future events, and to confirm the calculations by comparing them with actual observations or with records of observations. I refer of course to events that are not controlled by man. The events most useful for the purpose have been those the time of which can be calculated by the laws of celestial mechanics, most notably solar eclipses and occultations of stars by the moon. It should be noticed that when the observation of an event is used to verify our determinations of time (and intervals of time) we are making the time depend on the event and not the event on the time; in a sense the time becomes the dependent variable instead of the independent one. Of course the true state of affairs is that an equation expresses a relation between two or more variables; none of them is the independent variable in any absolute sense, but each is regarded sometimes as independent and sometimes as dependent, according to the nature of the operations being performed.

Returning to the verification of experiments, let us imagine a class of experiments in which a time interval plays an important part, as for example the measurement of velocities or accelerations, and in which a variable may be measured with a precision comparable to that with which the interval of time involved is known. Let us imagine further that two experiments performed at different times give inconsistent results. It is evident that the question arises whether it is the physical measurement or the time interval that is at fault. If it is possible to fit the time interval into its place in the whole scale of time, which is already known to be consistent with an important class of experiments (the observation of astronomical phenomena), that is, if we know the actual times between which the experiment was conducted, then the time interval may be said to be confirmed.

Our scale of time, then, together with the laws of celestial mechanics, is basically the means by which we know, both in advance and in retrospect, when astronomical phenomena may be observed. The comparison of a prediction with an observation is commonly called a comparison between theory and observation. Among the great number of sorts of astronomical phenomena that may be observed there are several discrepancies between theory and observation. In all such cases that I know of, the reason is simply that the calculations are so difficult as to be beyond our powers. The most

notable example is the rotation of the earth, which is not uniform, but is subject to accelerations of apparently random nature, such that if two different intervals of five years are chosen at random, the average duration of a day in one of them is likely to differ from the duration in the other by a few parts in 10^9 (Brouwer, 1952). Presumably the cause is to be found in the interior of the earth; its nature is not well understood.

It was stated that the measurement of time is basically the counting of the repetitions of a recurring phenomenon. In the case of certain phenomena the statement is clear enough; for example there is no difficulty in counting the ticks of an ordinary watch or the luminous periods of a flashing electric sign, and thus arriving at a measurement of time. Such measurements are not very useful, because they are not sufficiently long continued nor sufficiently accessible to all. What is required for a fundamental measurement of time is a phenomenon that can be observed by anyone who provides himself with the necessary equipment, and that will give results consistent with all observable phenomena. The phenomena actually used for measuring time are the revolutions of the earth around the sun. It is not obvious how revolutions of the earth are to be counted, and it is best to approach the subject in easy stages, by first considering how we may count rotations of the earth.

ROTATION OF THE EARTH

In a first approximation all that is necessary is to mount a telescope vertically, which can be done with the aid of a plumb bob, to provide a fine wire stretched north and south in the focal plane, and to note the instant when the image of any star as it passes through the field of view is quenched by the wire. The period of the rotation of the earth may be defined in a first approximation as the time interval between two successive passages of the same star. The purpose of the telescope is principally to constrain the observer's vision along a line of sight that is fixed with respect to the earth.

This crude procedure, if the observer is within say 40 degrees of the equator of the earth, gives a precision of about a part in 10^6, most of the error being made by the observer. By providing him with mechanical aids the error is reduced to a part in 10^7. In order to obtain still higher accuracy it is necessary to take account of the changes in the direction of the axis of rotation of the earth relative to the star. In the language of Newtonian mechanics, these changes are caused by the gravitational action of the moon and sun on the equatorial bulge of the earth, the moon and sun not moving in the plane of the equator of the earth. I shall pass over the operations by which the direction of the axis of rotation is calculated, because they do not differ in principle from the operations connected with the motions of planets around the sun, which will be treated at some length later. After the motion of the axis of rotation is allowed for, it is found that the next most important source of error is the accidental errors of observation. It is practicable, by observing many stars instead of one, to reduce the effect of the accidental errors by about a factor of 10, and thus the period of one rotation of the earth becomes measurable to a few parts in 10^8. At this stage the remaining error is not predominantly due to one cause, but to a combination of several. Among them are the individual motions of the stars relative to one another, variations in the direction of the axis of rotation relative to the observer, and the remaining accidental errors of observation. By means of statistical analyses of stellar motions and special series of observations the error may be still further reduced to a part in 10^8. Beyond this it does not seem necessary to go, because the variations in the period of rotation, mentioned earlier, set a working limit on the usefulness of the applications.

Before leaving the subject of the rotation of the earth, a few remarks about the frame of reference are in order. The frame of reference implied by what has already been said is the system of stars that are observed. Since each star is moving relative to the others, there is no reason to suppose that the mean motion is zero. It is more likely, *a priori*, that there is a small residual rotation of the stars as a system, and indeed it turns out to be so. Two methods are known for determining the residual rotation: the kinematic method and the dynamical one.

The kinematic method depends on extending the frame of reference outward from the earth to the most distant visible objects, the extragalactic nebulae (Wright, 1950). If the red shifts in nebular spectra are interpreted as velocities of recession, then the more distant of the nebulae that can be well observed are receding from us at very high velocities, more than 10 percent of the velocity of light. The velocities across the line of sight are entirely unknown. But if we assume that the cross-velocities are of the same order of magnitude as the velocity of recession, then, by reason of their very great distance, the *angular* velocity across the line of sight is negligibly small for many centuries. The stellar motions are so small that several decades elapse before they can be satisfactorily measured with respect to the nebulae; the work is now in progress. The indications are that the uncertainty in the present frame of reference sets a limit on the precision of time determinations such that the duration of a day cannot be measured with a precision greater than a part in 2×10^{15} at present. It would appear that this very high precision does not constitute an effective limitation on any experiments, but such a conclusion would be false. The effective precision with which the motion of the planets around the sun, or the moon around the earth, have been observed is about 0.1 second of arc per century, which corresponds precisely to the precision

with which the frame of reference is known. The observations of the present day are more accurate than those of a hundred years ago, and so an improvement in the frame of reference is desirable, at least in connection with studies of celestial motions. The dynamical method of establishing a frame of reference will be described later, in connection with the determination of time from planetary motions, a subject to which we now turn.

THE SECOND

For half a century astronomers have found that the rotation of the earth does not measure time with sufficient precision for their purposes. During much of this time there has been some confusion. There was natural reluctance to give up a measure of time that had been regarded as fundamental since time immemorial. Indeed, the day had become so firmly fixed in mind as an invariable unit of time that only a daring thinker could imagine any alternative. At last, however, the work of E. W. Brown (1926), W. de Sitter (1927), and H. Spencer Jones (1939) in succession proved beyond a doubt that it was necessary to relinquish the day in favor of a unit of time that would better correlate observable phenomena. By now the realization of the facts has become sufficiently general that the definition of the second has recently been changed by international authority. Whereas formerly the second was defined as 1/86 400 of a mean solar day, it is now 1/31 556 925.974 7 of the tropical year for 1900 January 0 at 12 h. ephemeris time.

The effect of the new definition is to make the second equal to the average of the old one during the 18th and 19th centuries. Whereas time reckoned in the old seconds was called mean solar time (universal time if reckoned from midnight on the meridian of Greenwich), time reckoned in the new seconds is called ephemeris time. The difference between the two has now accumulated to an amount such that a clock keeping ephemeris time would be about 30 seconds fast on a clock keeping universal time. The difference will not cause any confusion, because our ordinary clocks will continue to keep universal time as before, and when ephemeris time is desired, the necessary corrections will be applied in the calculations. Frequency standards, on the other hand, will be operated on the new time; inasmuch as the time of day cannot be obtained from a measurement of frequency, no confusion can possibly arise. An alternative possibility would have been actually to have altered our clocks to ephemeris time, but this procedure would have worked a hardship on navigators and surveyors, who must determine time in field operations in order to find their longitude; they have no means of determining ephemeris time, and it is better to throw the hardship on persons who have the means of coping with it.

There are two important differences between the new and old definitions of the second. The specification of a particular instant of time in the definition requires us to know at all times the relation between the present tropical year and the one for 1900.0. The other difference is that the tropical year, while it is in a sense the average duration of a revolution of the earth around the sun in a specified frame of reference, is not equal to the duration of any particular revolution; in fact one particular revolution may differ from another by as much as 20 minutes, or a part in 3×10^4. The mean solar day, on the other hand, is precisely equal to any rotation of the earth in its specified frame of reference.

REVOLUTION OF THE EARTH

Before considering what is actually counted in determining ephemeris time, let us pause to examine how a particular revolution of the earth may be observed without recourse to any theory. The annual revolution of the earth around the sun produces an apparent motion of the sun against the background of stars. I shall ask the reader to admit without proof that the angular position of the sun with respect to the stars in the nearby background can be ascertained with high precision at any instant; the observational techniques by which it is done are too complex to describe here, but they do not involve any new principle. If the apparent path of the sun among the stars is traced out with care for some years, it will be found that the path is very nearly a great circle on the apparent celestial sphere. The great circle is slowly rotating about one of its diameters, at the rate of about 47 seconds of arc in a century, and the maximum departure of the sun from the great circle is less than a second of arc. Both the rotation of the great circle and the departures of the sun from it are caused, again in the language of Newtonian mechanics, by the gravitational action of the other planets on the earth, Venus and Jupiter predominating. It is easy to choose a star not far from the axis of rotation of the great circle, and to count successive passages of the sun by this star. The frame of reference is the same as was used for counting rotations of the earth, and it may be refined and made more precise in the same way.

EQUATIONS OF MOTION

In the actual measurement of ephemeris time it is necessary to have recourse to the equations of planetary motion. Stated as concisely as possible, the method is as follows. The solution of the equations gives the angular position of the sun among the stars as a function of the ephemeris time. The solution is then inverted, so as to give the time as a function of the position of the sun. An observation of the sun is then used to establish the time at the instant of observation.

The equations of planetary motion employed are the equations of general relativity, but not in the form that they take most naturally. For numerical applications

it is easier to begin with the Newtonian form of the equations, and to express the difference between the two as a power series, the successive terms of which contain increasing powers of the velocity of light in the denominator. It is found in practice that only the first term of the power series is necessary in order to obtain a solution comparable in precision with the observations (Eddington, 1924; Chazy, 1928, 1930).

The equations of motion take many different forms according to the system of coordinates employed. Let us fix our attention on the equation for the orbital longitude, which is the angular distance of the sun, reckoned in the direction of motion from an origin to be described later, in the slowly rotating great circle before mentioned. Denoting the orbital longitude by L, we may put the solution of the equation in the form

$$L=C+Nt+St^2+E\sin[C-P+(N-R)t] \quad +\text{periodic terms.}$$

In this equation C, N, E, P are four constants of integration, which must be determined by observations of the sun. The constant C is the value of L, diminished by all periodic terms, at the fundamental epoch, say the beginning of the year 1900, from which the ephemeris time t is reckoned positively and negatively. The constant angle N is called the mean motion; evidently its numerical value depends on the unit in which t is reckoned. The unit of t might be chosen so as to make N unity, but it is preferable for practical reasons to connect t with the rotation of the earth, so that ephemeris time will not differ greatly from mean solar time. In the 19th century, when the speed of rotation of the earth was supposed to be constant, the unit of t was taken to be precisely 365.25 rotations, and N was determined by comparing the equation with observations during the 18th and 19th centuries. We still retain the same numerical value for N, whence it follows that the unit of t is 365.25 *average* rotations of the earth during the centuries mentioned, and N is nearly but not quite 2π radians. The remaining two constants of integration, E and P, do not require discussion. The constant S as it comes out in the calculations is about 10^{-11}; it is important to notice that although it has the form of an acceleration, it is not to be physically interpreted as such. Its presence is due merely to the form of the solution; other forms are known in which S vanishes, but those forms are associated with very difficult calculations, and they have not been used in practice. The periodic terms, a few hundred in number, are all of the form

$$k\sin(nt+K),$$

where k, n, and K are constants obtained during the course of the solution, k ranging up to about 10^{-4} and the periods varying between about a month and 1800 years. Finally, the constant R, according to Newtonian theory, is about 3×10^{-4}, which is increased a part in 1500 under general relativity. The small increase in R is the only difference between the solutions as given by Newtonian mechanics and by general relativity.

Having obtained the solution it is easy to make a table, or ephemeris, showing values of t and of L in parallel columns. An observation of L may then be used to determine t by interpolating in the table. It can now be seen that what is actually being counted, with the aid of the table, are units of t, which can be translated into ephemeris seconds using the new definition of the second already given.

So far nothing has been said to justify use of general relativity in preference to Newtonian mechanics. If observations of the sun were the only means of determining ephemeris time we should have no operational reason (so far as the measurement of time is concerned) for preferring one to the other. In fact, however, equations similar to the one already given may be calculated for the other planets, the moon, and the principal satellites of Jupiter, and the ephemeris time determined from observations of any of these objects. When this is done it is found that the observations of Mercury and Mars cannot possibly be reconciled with each other without introducing the modified R required by general relativity, or some modification of very nearly the same amount. The same conclusion is indicated, though with less certainty, by observations of Venus and the sun. The other objects are less affected by changes in R, or are observed with less accuracy, and no conclusions can be drawn from them.

FRAME OF REFERENCE

The origin from which L is reckoned may be established kinematically, as was described in connection with the rotation of the earth. But it may also be established dynamically, by comparisons of ephemeris time determined by different planets (Clemence, 1951). To understand how it is done let us imagine that our adopted frame of reference, including the origin from which L is reckoned, is slowly rotating about some axis. In general the rotation will have a component along the great circle in which L is measured, and this component will be absorbed in the constant of integration N, and also in $N-R$. A small change in $N-R$ produces in L a periodic oscillation with period $N-R$, phase a quarter-period different from that of the term with coefficient E, and amplitude proportional to t. Since the oscillation is present in L it is also present in the ephemeris time, which is determined from observations of L, but must remain undetected as long as the ephemeris time is determined from observations of the sun alone. Let it now be supposed that the ephemeris time is also determined by observing the L of another planet, which is reckoned in nearly the same plane as the L of the sun. The values of C, P, and N being different for the sun and the planet, the oscillation in the ephemeris time has a different phase and period, and if the ephemeris times

from the sun and planet are compared, there will be found to be an oscillating discrepancy between them. It will be found possible to remove the discrepancy by adjusting the two values of N by the same small amount, which is equivalent to altering the frame of reference previously adopted. Summing up the whole matter in a few words, the dynamical frame of reference is specified by the condition that observations of different celestial bodies shall yield the same ephemeris time.

Since the equations of motion give the values of the second derivatives of the coordinates with respect to the time, the periodic oscillations just referred to are the doubly integrated effects of periodic accelerations. So the dynamical frame of reference may also be defined operationally as that frame in which the planets exhibit no periodic accelerations not given by the equations of motion.

It may be instructive to inquire how the Newtonian frame of reference differs from the frame of general relativity, since it is not immediately evident how the modification of the N for different planets can be distinguished from the modification of the Newtonian value of R required by general relativity. The answer is that different modifications of R are required for the different planets, these modifications being given by general relativity itself, whereas one and the same modification of N is required for all the planets, this modification being given by a comparison between observation and theory. Thus, if it is attempted to establish a dynamical frame of reference using the Newtonian equations of motion, it will be found that a different frame is required for every planet.

There is nothing in the operations that obviously requires the dynamical frame of general relativity to be identical with the kinematical frame discussed earlier. It turns out experimentally that if there is any rotation of one with respect to the other, the integrated effect during a century is less than 0.1 second of arc. The precision of observations and the difficulty of the calculations do not permit us to say more at present, but it is expected within another decade or two to reach another decimal of a second.

Throughout the discussion of the determination of ephemeris time I have been careful to avoid any mention of measurement of lengths, in order to show that it need not be introduced. Indeed it need not. Rectangular coordinates are introduced only as a computational convenience. Ephemeris time could in fact be determined by observations of the sun and moon alone, completely neglecting the circumstance that they are moving during the interval while light is coming to us from them, the effects of the motions being absorbed in the constants of integration for the most part. Residual errors would appear as periodic discrepancies between observation and theory, which would be too small to be detected with certainty. For the general correlation of astronomical phenomena it is, however, necessary to make allowance for the finite velocity of light, and for convenience it is done in all of our theory. Since the procedure is not germane to the determination of time I shall not describe it here. It is pertinent to remark only that the distances are not measured directly but inferred from the angular velocities, that the unit of distance is completely arbitrary, and that the interval of time required for light to traverse the unit of distance may be determined with sufficient precision (three or four significant figures) from observations of the phenomena themselves.

CLOCKS

Let us turn our attention to the definition of a clock. We leave aside the subject of atomic clocks, which although important, is outside the scope of this article, and confine attention to astronomical clocks. Evidently in common practice a clock is a mechanism for measuring time that is approximately synchronized with mean solar time, that is, with the rotation of the earth. But this definition is not suitable for precise applications, because as we have seen, the rotation of the earth is subject to unpredictable variations. It seems to me that a clock must be defined as a mechanism for measuring time that is continually synchronized as nearly as may be with ephemeris time. This definition implies an observer who initiates the synchronization, and the observer must be located on the earth, since ephemeris time is defined only for an observer on the earth. We may imagine the clock to be transported anywhere we like, either on or off the earth, but we must always imagine as well that there is an observer who remains on the earth and synchronizes the clock with ephemeris time, either by means of light signals, or in some other specified manner. Perhaps it may be possible to extend the definition of ephemeris time, so as to render it determinable by an observer not on the earth; the definition would then lose something of its operational quality, because the operations described would not be those actually performed, but those that are imagined to be performed, which is not quite the same thing.

PRECISION OF TIME INTERVALS

The relative precision with which an interval of ephemeris time can be determined depends partly on the precision of the man-made clocks that are used to interpolate between astronomical observations and to smooth out to some extent the accidental errors of the observations, partly on the precision of the astronomical observations, and partly on the time itself, since the precision of clocks and of observations have been continually increasing for several centuries. Also it is necessary to distinguish between the precision attained in ordinary practice, and that attained when an extraordinary effort is made. Furthermore, a time interval can always be stated more precisely some time after it is

past than it can be immediately after the measurement is made, because interpolations of astronomical observations are more precise than extrapolations; for this reason the principal time services of the world issue bulletins, which are always some weeks or months in arrear, giving revised estimates of the times at which their radio signals were propagated, and the Bureau International de l'Heure issues, a year or so later, still more accurate estimates. In view of all these considerations no statement about the precision of time intervals can be very simple and at the same time very meaningful. It is nevertheless possible to give some understanding of the subject in a few words.

Intervals of time varying from very short ones, say 10^{-9} second, to those of moderate length, say 10^6 seconds, can be determined with moderate effort to a precision of a part in 10^8; longer intervals, of about 10^7 seconds, to a precision of a part in 10^9; and still longer ones, of 10^8 seconds or a little longer, to a precision of a part in 10^{10}. The longest interval that can be said to have been actually measured is the interval from some ancient eclipses of the sun to the present time, about 10^{11} seconds, which is probably known with a precision of a part in 10^8; the uncertainty is mainly in the interpretation of the ancient records.

Dirk Brouwer, Astron. J. **57**, 125 (1952).
E. W. Brown, Trans. Astron. Obs. Yale Univ. **3**, 205 (1926).
Jean Chazy, *La Théorie de la Relativité et la Mécanique Céleste* (Gauthier-Villars et Cie, Paris), Vol. I (1928), Vol. II (1930).
G. M. Clemence, Monthly Notices Roy. Astron. Soc. **111**, 220 (1951); Am. Scientist **40**, 260 (1952); Science **123**, 567 (1956).
A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, New York, 1924).
H. Spencer Jones, Monthly Notices Roy. Astron. Soc. **99**, 541 (1939).
W. de Sitter, Bull. Astron. Inst. Netherlands **4**, 21, 49, 69 (1927).
W. H. Wright, Proc. Am. Phil. Soc. **94**, 1 (1950).

Shifting of the Earth's Axis of Rotation*

DAVID R. INGLIS
Argonne National Laboratory, Lemont, Illinois

The terrestrial variation of the earth's axis of rotation includes two closely related phenomena, the Euler-Chandler nutation and the slow wandering of the pole. These are quite separate from the 27 000-year precession in space due to external torques which are here ignored. In Part I an attempt is made to protray in simple terms the effect of the elastic deformation which takes place in the free nutation and is held responsible for the slowing down of the 306-day Euler nutation of a rigid earth to the observed Chandler period of about 420 days. A simplification is given of the Klein-Sommerfeld derivation of this old result of Hough which is based on assuming complete elastic readjustment to the instantaneous axis of rotation. For further elucidation a model is introduced which treats the axial and the equatorial regions as coupled gyroscopes. The model is modified to permit a viscous slipping of the elastic coupling, and in terms of this the damping of the nutation is discussed. In Part II this model is further modified to facilitate discussion of the mechanism proposed by Gold to explain polar wandering, depending on static asymmetric deviations from a perfect figure of the earth, such as excess mountains in two opposite octants. On the basis of the rate of viscous readjustment as given by the damping of the nutation, it is shown that the possible asymmetries to be expected from the existence of continents in isostatic equilibrium are sufficient to give a rate of wandering much more rapid than needed to explain the long-term motion of the pole in geologic times, in keeping with Gold's suggestion that it may have taken place in a sequence of relatively short spurts from one temporarily stable position to another. The melting of circumpolar ice provides a feedback mechanism to stabilize the present position. Fermi suggested that the angular momentum of marine currents might be sufficient to explain the slow polar wandering. It is shown that this is indeed sufficient but a much less important contribution to the balance than the continental asymmetries.

INTRODUCTION

NOT only does the earth's axis of rotation sweep through the heavens in its familiar 27 000-year precession arising from the couple applied by the second derivative of the inverse-square forces of the sun and moon acting on the earth's quadrupole moment, but it also varies slowly relative to the earth itself. This terrestrial variation of the pole consists of a somewhat irregular nutation and a slower wandering that is significant in geological times. The nutation has a period of a little over a year and very small amplitude, remaining within about a second of arc, or a hundred meters, of a mean position of the pole. There is paleontological evidence, such as the distribution of coal deposits, to indicate that the pole has wandered through large angles in the last 10^8 years, having once been in the vicinity of Hawaii. The recent development of paleomagnetism holds promise of giving further detail. Doubts of earlier physicists that the pole could wander nurtured discussion of such implausibly energetic phenomena as distant migration of continents during these times, and it is gratifying that recent evidence concerning the elasticity and plasticity of the earth supports the possibility of a much gentler wandering of the pole.[1] The tumultous history of the subject includes prominently the names Darwin and Kelvin, as has been reviewed recently by Munk.[2] The nutation and polar wandering are related by being separate manifestations of the earth as an almost symmetrical and not quite rigid top. The interpretation of the nutation is used to calibrate the plasticity for use in understanding the wandering.

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] T. Gold, Nature **175**, 526 (1955).
[2] W. H. Munk, Nature **177**, 551 (1956).

PART I. THE FREE NUTATION OF THE ELASTIC EARTH

THE EQUILIBRIUM-DEFORMATION MODEL

A complete treatment of the motion of nutation of an elastic spheroid in free space would involve consideration of the gyroscopic effects contributed by the detailed displacements of all elements of volume reacting to the varying centrifugal and Coriolis forces. Such a treatment would be very formidable and has not been carried out. The earth deviates very little from spherical shape, its ellipticity (effective ellipticity as indicated by the quadrupole moment in the 27 000-year precession) being about 1/306, with the result that the pole moves around its average geographic position with a period of several hundred days. The small ellipticity and slow nutation conspire to pick out certain aspects of the deformation, essentially the static-equilibrium aspects, as dominant and relegate some of the dynamic details to higher order in the small ellipticity. This leads to the greatly simplified model in terms of which Klein and Sommerfeld[3] have shown that the period of nutation should be considerably greater than the Euler period of 306 days expected for a rigid earth, in qualitative agreement with the observed Chandler period of about 427 days.

[3] F. Klein and A. Sommerfeld, *Theorie des Kreisels* (B. G. Teubner, Leipzig, 1903), Vol. III. Other early treatments of the problem were given by (a) S. Newcomb, Monthly Notices Roy. Astron. Soc. (1892); (b) S. S. Hough, Trans. Roy. Soc. (London) A, **187** (1896); (c) G. Herglotz, Z. Math. Physik **52**, 275 (1905); (d) A. E. H. Love, Proc. Roy. Soc. (London) **A82**, 73 (1909).

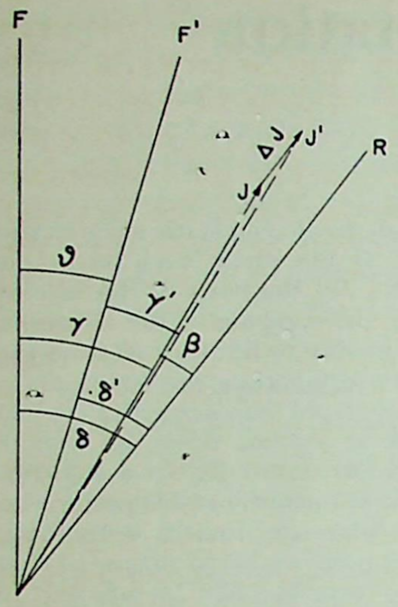


FIG. 1. Angles in the Klein-Sommerfeld model. The notation follows reference 1. R is the axis of rotation. F and J are the principal axis and angular momentum of a body having the shape to which the earth would relax elastically if it were not rotating. F' and J' refer to the actual earth, stretched from that shape by centrifugal force.

Although the derivation given by Klein and Sommerfeld is rather lengthy, covering 30 pages of their book (pages 598 to 608 and 685 to 706), the essentials of it can be explained rather briefly in a reasonably graphic manner. They discuss explicitly a spheroid having a small static ellipticity ε, to which is added another small ellipticity ε' arising from the elastic deformation resulting from centrifugal force, and they apply this only at the end to an earth in which the shape during rotation is taken to be the equilibrium shape of no internal strain. The distinction is almost trivial, but we shall here consider explicitly the latter, more geophysical, case. The essential assumptions of the model they use are that the additional elastic reaction to centrifugal force is an equilibrium deformation, accomplished instantly to fit the axis of rotation, and that the moments of inertia resulting from these superposed ellipticities completely characterize the features of the body which affect the rotation.

We first consider the earth rotated with an angular speed Ω about an axis F (z axis) for a time long enough to establish an equilibrium figure without internal strain, through the mechanism of slow plastic flow, the ellipticity

$$E=(R_{\max}-R_{\min})/R_0=(I_z-I_x)/I_x \qquad (1)$$

being determined by a balance between gravity and centrifugal force, as for a fluid. (Here I_z and I_x are principal moments of inertia of the spheroid and R is its radius as a function of latitude θ.) At the end of our derivation, F will turn out to be the average position of the axis of figure, to which the slow plastic readjustment will naturally be made. If the rotation were stopped for too short a time for plastic flow, the axis of figure would remain F but the ellipticity would be

$$\varepsilon=E-\varepsilon', \qquad (2)$$

where ε' is the ellipticity which a spherical elastic earth would acquire if rotated at an angular speed Ω, the neglected difference between the deformation given by the rotation to a spherical earth and to an elliptical earth being of higher order in these small ellipticities. If a fictitious rigid earth of this shape (2) were rotated with angular velocity $\mathbf{\Omega}$ about an axis R making a small angle δ with F, it would have an angular momentum $\mathbf{J}$ between R and F, as in Fig. 1, making a much smaller angle β with R because the moment of inertia C about F is only slightly larger than the moment of inertia A about an axis normal to F. [In terms of these, $\varepsilon=(C-A)/A$ (see Eq. (3) below).] The actual elastic earth rotated with $\mathbf{\Omega}$ about R has a shape which differs from this by the added deformation corresponding to an ellipticity ε' and an axis of symmetry R. The superposition of two spheroidal displacements from spherical shape, one of ellipticity ε with axis F and one of ellipticity ε' with axis R very near F, gives a spheroid of ellipticity $E=\varepsilon+\varepsilon'$, as in (2), because near a principal axis the radius of an ellipse as a function of the angle has no first-power term in the angle. F' lies between F and R because the figure is a compromise between that for a fluid earth, for which F' would be along R, and for a rigid earth, for which F' would be along F. [If one cares to be quantitative about this point, the radius $r(\theta)$ can be expressed as a superposition of the effects of ε and ε' as in Eq. (5) on page 602 of reference 3 and the angle

$$\vartheta=(\varepsilon'/(\varepsilon+\varepsilon'))\delta$$

between F' and F may be obtained from $dr/d\theta=0$, expanded for small δ.]

The angular momentum of the actual earth rotating with $\mathbf{\Omega}$ about R is J'. The important point of the whole derivation, which may be seen very easily vectorially although it is hidden in a string of formulas in reference 3, is that J' is very nearly parallel to J, the difference in their direction being of higher order in the small quantities that any of the other relevant angles. The reason for this is that $\mathbf{\Delta J}$, the difference between $\mathbf{J}'$ and $\mathbf{J}$, is the angular momentum contributed by the ellipticity ε' arising elastically from the rotation of the elastic sphere about R, and is therefore parallel to R, which means very nearly parallel to $\mathbf{J}$. Moreover, $|\mathbf{\Delta J}|$ is small compared with J because ε' is small.

Aside from this, all that is needed is the application of the ideas that lead to the simple Eulerian nutation twice, once to the fictitious rigid earth with ellipticity ε' (to which the earth would relax elastically if it should stop rotating) and once in a slightly modified manner to the actual earth. The angle between $\mathbf{J}$ and R (see Fig. 2) is obtained as

$$\beta=\delta-\gamma=\Omega_x/\Omega_z-J_x/J_z=(\Omega_x/\Omega_z)(1-A/C)=\delta\varepsilon \qquad (3)$$

in the limit for small angles equal to their sin and tan. For the usual reasons given in mechanics texts[4] these angles are constant, and the motion of the body in space

[4] For example, D. R. Inglis, *Dynamic Principles of Mechanics* (The Blakiston Company, New York, and McGraw-Hill Book Company, Inc., New York, 1949), pp. 136 ff.

is given by the rolling of two cones on one another in contact along the instantaneous axis of rotation R, a fairly narrow body-fixed cone with axis F on a very narrow space-fixed cone with axis **J**. Seen from the earth, the space-fixed cone rolls in the larger one, the end of the vector **J** moving at a rate $\Omega\beta J$ about a circle which it also moves at the rate $\Omega_n\gamma J$, where Ω_n is the angular speed of the nutation of this fictitious right earth in the terrestrial system,

$$\Omega_n = (\beta/\gamma)\Omega = (\delta/\gamma)\varepsilon\Omega = \varepsilon\Omega, \tag{4}$$

the familiar Euler result.

In the actual earth rotating with the same $\mathbf{\Omega}$, we have a different figure axis F' making an angle ϑ with the former F, new angles δ' and γ' but $\beta' = \delta' - \gamma'$ equal to the former β, because J' is parallel to J to a sufficient approximation, as we have seen. Because the earth is an oblate spheroid, J' remains between R and F', and the next increment of motion is a rolling of a space fixed surface on a body-fixed surface, both normal to the plane of R and F'. The special assumption of the model is that there is no lag in the elastic deformation arising from centrifugal force, or that F' remains in the plane of R and F. The body-fixed surface is then normal to the plane of R and F, and F is fixed in the body, so the surface is a (circular) cone with F as its axis, *just as for the fictitious rigid earth*. The space-fixed cone now has $\mathbf{J}'$ as its axis, so the nutation is described by the rolling of the same two cones on one another as before, and the angular velocity of nutation is again Ω_n, given by (4).

The result, originally due to Hough,[3b] is thus that the period of nutation is the same as it would be if the earth were rigid and had the smaller ellipticity to which the actual earth would relax elastically if the rotation were stopped. If it were rigid with its actual ellipticity, it would have an Euler period of about 306 days. It has a period of about 427 days (in a nutation with a varying amplitude). It would have a rigid rotation with this period of nutation (and constant amplitude) if it should relax elastically to an ellipticity $306/427 = 0.72$ as great as it has. Klein and Sommerfeld (page 701) have shown that this would require it to be very stiff, with a coefficient of rigidity (shear modulus of elasticity) μ about 24% greater than that of steel (which is 2×10^{11} dynes/cm^2). This estimate is based on the assumption of an incompressible homogeneous earth. The fact that the earth has a fluid core occupying more than the inner half of the radius makes it considerably easier to distort than it would be if the core were solid, so suggests that the mantle would have to be much stiffer than steel, as in fact it seems to be. Evidence from the velocity of shear waves indicates that the rigidity increases rapidly with pressure, the coefficient of rigidity averaging about 2×10^{12} dynes/cm^2, ten times that of steel, in the mantle.[5] This doubtless arises from the difficulty of sliding layers of atoms in a close-packed lattice under high pressure. Takeuchi[6] has shown that this high rigidity of the mantle is compatible with observed earth tides, if the core is fluid or indeed has any rigidity up to about that of steel. It thus remains plausible within the framework of the equilibrium-deformation model that the discrepancy between the Chandler period and the Euler period is to be ascribed to the elasticity of the earth.

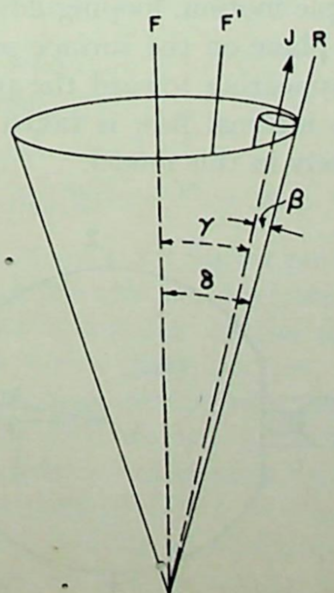


FIG. 2. Rolling-cone construction for the Eulerian nutation.

SUMMARY OF MECHANISM

Let us recapitulate the mechanism of the equilibrium-deformation model in simple terms, and in words rather than formulae. When the earth rotates for a long time about one axis, this becomes the axis of symmetry and there is no internal strain after slow viscous flow. If the rotation were stopped for a short time, we would have a residual spheroidal distortion with the same axis of symmetry (with internal strain after elastic relaxation of the shape). The fictitious rigid earth with this distortion, if rotated about a nearby axis, has a rate of nutation determined by its two moments of inertia. An elastic earth rotating about the nearby axis differs from this by a centrifugal deformation which is symmetrical about that axis, and which alone would cause no nutation. Since these small deformations and their effects may be superposed linearly, the elastic earth is left with the nutation of the fictitious rigid earth. The smaller the deformation of a rigid top, the slower is its nutation. The earth thus has a slower nutation than it would have if it were rigid with its actual, larger, deformation.

VARYING AMPLITUDE OF THE NUTATION

In view of the complexity and the possible annual variations of the earth's aqueous and atmospheric overburden, it is no wonder that so small an effect as the nutation progresses in a highly irregular manner. The

[5] B. Gutenberg, *International Constitution of the Earth* (Dover Publications, New York, 1951), p. 369.

[6] H. Takeuchi, Trans. Am. Geophys. Union **31**, 651 (1950).

main feature of the irregularity appears to be a variation of the amplitude of the nutation with a period of about nine years. This has been tentatively interpreted as a beat frequency between the Chandler period and the annual forcing term arising from seasonal variations of the overburden. Without entering here into the complicated discussion of these effects, it is worth pointing out how the question of the validity of the assumptions that the elastic deformation follows instantaneously the change of the axis of rotation and that there is no plastic flow are relevant to this problem.

The assumption of instantaneous readjustment of the figure to the axis is justified by the fact that an elastic effect should take place in a time determined by the transit of elastic waves across the earth, which is leas than an hour, very much less than the 427-day period of the nutation. If there were nevertheless a lag of this sort it would make F' lag behind the F-R plane, and instead of the circular cross section of the cone about F we would have a gradual inward spiral, corresponding to a damping of the nutation. An effect similar to this would also result from slight plastic flow within the nutation period.[7] In this case F, rather than being fixed in the body, would be relaxing slowly toward R at any instant, again making R move along an inward spiral about F in a damped nutation. It seems plausible that this damping, which is just one of the competing effects, should not be more rapid than the change of the varying amplitude, as has been assumed in getting an indication of the plastic flow within the earth.[3,8] The latest determination[9] suggests that the rate of decay of δ in the nutation is 3.1 Chandler periods or about four years, although values up to ten years[8] are still perhaps not to be excluded because of indefiniteness of the mechanism by which the nutation is excited—it includes a considerable contribution from seasonal changes in the products of inertia due to snow accumulation on continents, for example, but apparently some random changes also.

NEGLIGIBILITY OF DYNAMIC EFFECTS OF THE RATE OF DEFORMATION

To justify the equilibrium-deformation model for small ellipticities, it is instructive to estimate the transverse angular momentum contributed by the volume flow incidental to the variation of orientation of the deformation. The equatorial bulge above the surface of a spherical earth has a mass of the order $\varepsilon\rho R_0^3$, or εM, where M is the mass of the earth. During half of the nutation period it moves along a meridian a distance δR_0, taking a time $1/\varepsilon\Omega$, and thus having a velocity $\varepsilon\delta\Omega R_0$ and an angular momentum in the equatorial plane about equal to

$$\varepsilon^2\delta\Omega M R_0^2. \tag{5}$$

This is to be compared with the component of angular momentum normal to $\mathbf{\Omega}$ contributed by the difference $(C-A)$, which is

$$J_\perp = \beta J \approx \varepsilon\delta\Omega M R_0^2. \tag{6}$$

Thus the transverse momentum contributed by the distortion kinematics is smaller than that we have taken into account by a factor of the order of ε, and must be considered negligible.

A DOUBLE-GYROSCOPE MODEL OF THE NUTATION

For the sake of further understanding of the nutation mechanism, it is of some interest to consider an independent treatment by means of another model, in which the internal kinematic effects at first appear explicitly before they are neglected. This double-gyroscope model consists of a rigid central body having as principal axes a body-fixed coordinate system xyz with origin at the center of mass, and a ring mounted concentrically on it by fairly stiff springs so that its equilibrium position is in the equatorial xy plane as sketched in Fig. 3. The central body has principal moments of inertia A_0, B_0, C_0, with $A_0 = B_0$, and the ring likewise a, a, c. We consider the free rotation of this system about axes near the z axis. This model has some similarity to the problem of the earth's free rotation in space. In the limit of very stiff springs, it becomes a rigid symmetrical top with its moment of inertia $C = C_0 + c$ about z assumed to be slightly greater than those A about x and y. With some elasticity of the springs, the matter near the equator, in the ring, is free to shift slightly relative to the matter near the poles, as is true also of the matter near the equatorial "bulge" that accounts for the difference $C-A$ in the actual earth. The actual lines of flow are expected to be quite different from this simple motion, looping down beneath the surface from the plane on the surface where the equatorial bulge is disappearing toward the place where it is developing. The internal flow is taken into account only approximately in this model.

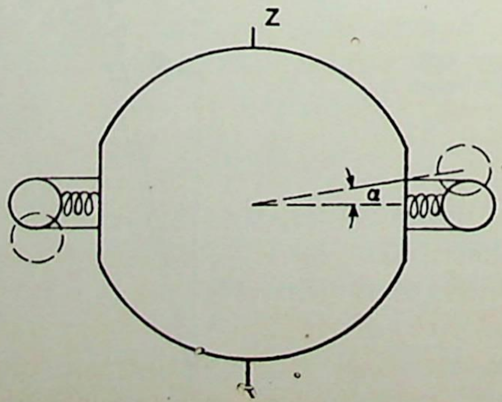


FIG. 3. Double-gyroscope model.

[7] J. M. Burgers has recently considered this question as a part of an elegant general treatment of the pole's shifting which includes the elastic nutation here discussed as a special case. Proc. Konikl. Ned. Akad. Wetenshap. B, LVIII, 219 (1955).

[8] H. Jeffreys, *The Earth* (Cambridge University Press, New York, 1952), p. 209, for example.

[9] A. M. Walker and A. Young, Monthly Notices Roy. Astron. Soc. 115, 443 (1955).

The angular displacement of the ring relative to the sphere, from the equilibrium position, is given by the vector $\boldsymbol{\alpha}$, and we assume that a constraint is provided (a key in a meridional slot) so that $\alpha_z=0$. The stiffness constant of the angular spring we call K, so the torque exerted on the sphere by the collection of springs is $K\boldsymbol{\alpha}$. The rates of change of the angular momenta $\mathbf{L}$ of the sphere and $\mathbf{l}$ of the ring are then

$$d\mathbf{L}/dt=-(d\mathbf{l}/dt)=K\boldsymbol{\alpha}+\lambda\mathbf{k}. \quad (7)$$

Here λ represents the unknown torque about the z axis exerted by the constraint. If we assume that the components α_x and α_y remain small compared to unity, the nonvanishing moments and products of inertia of the ring in the xyz coordinate system may be written

$$I_{xx}=I_{yy}=a, \qquad I_{zz}=c,$$
$$I_{xz}=\alpha_y(c-a), \quad I_{yz}=\alpha_x(a-c), \quad (8)$$

where we neglect squares and higher powers of the α's in applying the formula,

$$I_{ab}=\sum_{\alpha'}(i_a\cdot i_{\alpha'})I_{\alpha'\alpha'}(i_{\alpha'}\cdot i_b), \quad (9)$$

transforming by an angle α from the principal axes of the ring, in which $I_{x'x'}=I_{y'y'}=a$, $I_{z'z'}=c$. In the case of a thin ring, (8) and some of the following equations may be written a little more simply with $c=2a$, $(c-a)=a$. If we call the angular velocity of the central body in space $\boldsymbol{\Omega}$ and of the ring $\boldsymbol{\omega}=\boldsymbol{\Omega}+(d\boldsymbol{\alpha}/dt)$, the angular momenta of the central body and of the ring are, respectively,

$$\mathbf{L}=A_0(\Omega_x\mathbf{i}+\Omega_y\mathbf{j})+C_0\Omega_z\mathbf{k} \quad (10)$$

$$\mathbf{l}=\sum \mathbf{i}_\alpha I_{\alpha\beta}\omega_\beta=[a\omega_x+(c-a)\alpha_y\omega_z]\mathbf{i}+[a\omega_y-(c-a)\alpha_y\omega_z]\mathbf{j}+[c\omega_z+(c-a)(\alpha_y\omega_x-\alpha_x\omega_y)]\mathbf{k}.$$

By differentiation and use of (7) we obtain the equations of motion

$$[A_0(d\Omega_x/dt)+(C_0-A_0)\Omega_y\Omega_z]\mathbf{i}+[A_0(d\Omega_y/dt)-(C_0-A_0)\Omega_x\Omega_z\mathbf{j}=K\boldsymbol{\alpha}+\lambda\mathbf{k} \quad (11)$$

$$\{a\dot{\omega}_x+\Omega_z[(c-a)(\dot{\alpha}_y+\alpha_x\Omega_z)-a\omega_y]+\Omega_y[c\Omega_z+(c-a)(\alpha_y\Omega_x-\alpha_x\Omega_y)]\}\mathbf{i}+\{a\dot{\omega}_y-\Omega_z[(c-a)(\dot{\alpha}_x-\alpha_y\Omega_z)+a\omega_x]-\Omega_x[c\Omega_z+(c-a)(\alpha_y\Omega_x-\alpha_x\Omega_y)]\}\mathbf{j}=-K\boldsymbol{\alpha}-\lambda\mathbf{k}. \quad (12)$$

We have again neglected the second and higher powers of the α's, and in the vector differentiation we have, of course, used $d\mathbf{i}_a/dt=\boldsymbol{\Omega}\times\mathbf{i}_a$ as is familiar in the derivation of Euler's equations for a rigid rotator, for example. The terms $\alpha_y\dot{\omega}_z$ and $\alpha_x\dot{\omega}_z$ which one might expect from differentiation of (10) do not appear in (12) because $\dot{\omega}_z$ is small of order α, as is seen by examination of the third component of (7), $\dot{L}_z=-\dot{l}_z$, in the light of (10) and with the condition $\dot{\omega}_z=(d\Omega_z/dt)$.

As a point of departure for an approximation procedure, we first consider the limiting case in which the spring is rigid, $K=\infty$, $\boldsymbol{\alpha}=0$, $\omega=\Omega$. Then on eliminating $K\alpha$ from (11) and (12) we have the Euler equations

$$A(d\Omega_x/dt)+(C-A)\Omega_x\Omega_z=0$$
$$A(d\Omega_y/dt)+(A-C)\Omega_x\Omega_z=0, \quad (13)$$

where $A=A_0+a$ and $C=C_0+c$, the principal moments of inertia of the combined central body and ring, and Ω_z is constant. The trial solution, $\Omega_x=D\cos ft$, $\Omega_y=D\sin ft$, leads immediately to the well-known angular frequency of nutation $f=\Omega_z(C-A)/A$ in the body-fixed (earth's) coordinate system.

In the case of interest with K large and α small, we try a similar solution with a slightly different frequency $F=f+\epsilon$:

$$\Omega_x=D\cos Ft \quad \alpha_x=\alpha\sin Ft \quad \dot{\alpha}_x=\alpha F\cos Ft$$
$$\Omega_y=D\sin Ft \quad \alpha_y=-\alpha\cos Ft \quad \dot{\alpha}_y=\alpha F\sin Ft. \quad (14)$$

Because $F\approx f\ll\Omega_z$ and because Ω_x and Ω_y are much less than Ω_z, the terms $\alpha_x\Omega_z^2$ and $\alpha_y\Omega_z^2$ are much larger than the other terms in α on the left side of (12). It is the smallness of $F\alpha_y\Omega_z$ compared to $\alpha_x\Omega_z^2$, for example, that corresponds to the negligibility of dynamic effects of the rate of deformation in our earlier discussion. Neglecting those smaller terms, we have as the $\mathbf{i}$ component of (12), when combined with (11), the simpler equation

$$Ad\Omega_x/dt+(C-A)\Omega_y\Omega_z+(c-a)\alpha_x\Omega_z^2=0. \quad (15)$$

From either this or the similar equation for the j component, with (14), we have the equation

$$A(f+\epsilon)-(C-A)\Omega_z-(c-a)(\alpha/D)\Omega_z^2=0. \quad (16)$$

Here f is defined to cancel the middle term, that is, by the equality when the small terms vanish, and with

$$(\alpha/D)=[(C_0-A_0)\Omega_z-A_0f]/K \quad (17)$$

from (11), the small terms of (16) give

$$\epsilon/f=((c-a)/(C-A))(\alpha/D)\Omega_z=[(C_0-A_0)/(C-A)-A_0/A]\Omega_z^2/[K/(c-a)]. \quad (18)$$

The denominator $[K/(c-a)]$ is the square of a frequency, ω_0^2, and in the case of a thin ring with $c-a=a$ this ω_0 is the (angular) frequency with which the ring would vibrate tortionally about an axis in the xy plane if the central body were held fixed. If, further, the central body is a sphere, the term in (C_0-A_0) vanishes and the right side of (12) becomes $-\Omega_z^2/(K/\mu_0)$, where $\mu_0=A_0a/(A_0+a)$, the "reduced moment of inertia" for such vibrations with the sphere free to vibrate, analogous to the familiar "reduced mass" in the vibration of a model of a diatomic molecule.

For the sake of making a numerical estimate of the orders of magnitude involved, let us for a moment ignore the fluidity of the core of the earth, and consider the central body to be a cylinder about the z axis like the core of an apple, with a diameter about six tenths

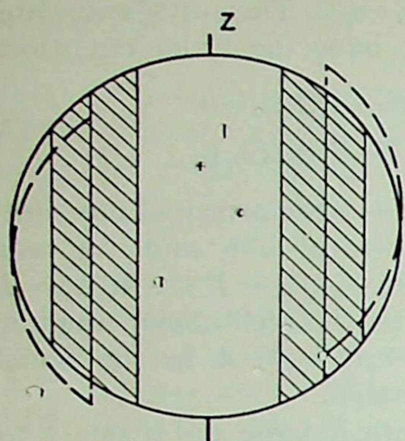


FIG. 4. Schematic representation of strain in the double-gyroscope model.

of the earth's diameter, surrounded by a thick annular ring representing the rest of the earth. With this division of a homogeneous earth, the outer ring and the cylinder each contain half of the earth's mass. We consider that shear takes place in a zone about half an earth radius thick, on both sides of the cylindrical surface (Fig. 4). The values of the moments of inertia are $(a,c,A_0,C_0) = (0.218, 0.311, 0.182, 0.090)MR_0^2$, respectively. (Since the model is a rough approximation at best, it would be sufficiently accurate to avoid the integrations here involved and instead take the rough values obtained by treating the central body as a cylinder, using the simple formulas $C_0=\frac{1}{2}(M/2)(0.6R_0)^2$ and $A_0=C_0/2+(M/6)R_0^2$, and subtracting these from the spherical value $(2/5)MR_0^2$ to get c and a.) From the observed gravitational quadrupole moment we take $C-A=(0.4/306)MR_0^2$, and further use a density 5.5 g/cm³, the observed mean density of the earth, and $R_0=6.3\ 10^8$ cm.

The motion which in Fig. 3 we considered a tipping of the ring relative to the central body here involves a more complicated shearing strain. If the part of the ring on the right side of Fig. 4 slides up by an amount dz while the opposite part slides down, we see only a simple shearing strain in the shaded regions in the figure, but in the planes above and beneath the figure there are twisting strains both in the ring and the intermediate region. As a rough approximation we assume that the shearing strain in the shaded regions, diminished by a factor $\sin^2\theta$ in the other azimuths, contributes the main part of the restoring force, as it does with the simpler geometry of Fig. 3 when the space between the thin ring and central body is filled up with rubber. The torque is then

$$T=K\alpha=Kdz/(0.6R_0)=(0.6R_0)(\text{area}) \times(dz/(R_0/2))(\sin^2\theta)_{\text{Av}}\mu, \quad (19)$$

relating K to the shear modulus of elasticity μ, with $(\sin^2\theta)_{\text{Av}}=\frac{1}{2}$, $(\text{area})=4\pi(0.6R_0)(0.8R_0)$.

From the fact that the quadrupole moment leads one to expect a 306-day nutation, compared with the observed period of about 427 days, we have

$$\epsilon/f=[1/427-1/306]/(1/306)=-121/427. \quad (20)$$

These various quantities in Eq. (18), along with $\Omega_z=(0.86\ 10^5\ \text{sec})^{-1}$, give as an estimate of the rigidity of the earth necessary to explain the nutation discrepancy

$$\mu=1.4\ 10^{10}\ \text{dynes/cm}^2 \quad (21)$$

less than one-tenth of the rigidity calculated by the more general Klein-Sommerfeld treatment of a homogeneous earth. By limiting the distortion to shear along planes normal to the undisturbed centrifugal force, we have severely reduced the freedom of distortion and thus a smaller μ is required to permit a given reaction to a given tilting of the axis than with more general distortions permitted. It is perhaps a little surprising that this limitation should have reduced the estimated μ as much as it appears to have done.

If, contrary to expectation for the earth, the central body were assumed to be a rigid sphere with a narrow ring more like that in Fig. 3 only heavy enough to account for $C-A$, the term in A_0/A that is left in Eq. (18) is less than a hundredth as large as the other term, and a is reduced by more than a factor 100, so the estimated μ is of the order of 10^{-4} smaller than in (21), its value depending on the geometry assumed for the region of shear.

The model as sketched in Fig. 4 is, of course, only intended to introduce the inertial properties of the various parts of the earth, and their elastic interaction, in an approximate way to bring out the main features of the effect as simply as possible. In the actual earth with its fluid core occupying more than the inner half of the radius, the actual shear lines will be quite different, but the model should still give the right order of magnitude. Although the fact that a sphere is a nondevelopable surface gives it more strength than is apparent in a sectional figure, there takes place against this stiffness a displacement by the bending of the mantle, as indicated schematically in Fig. 5, in addition

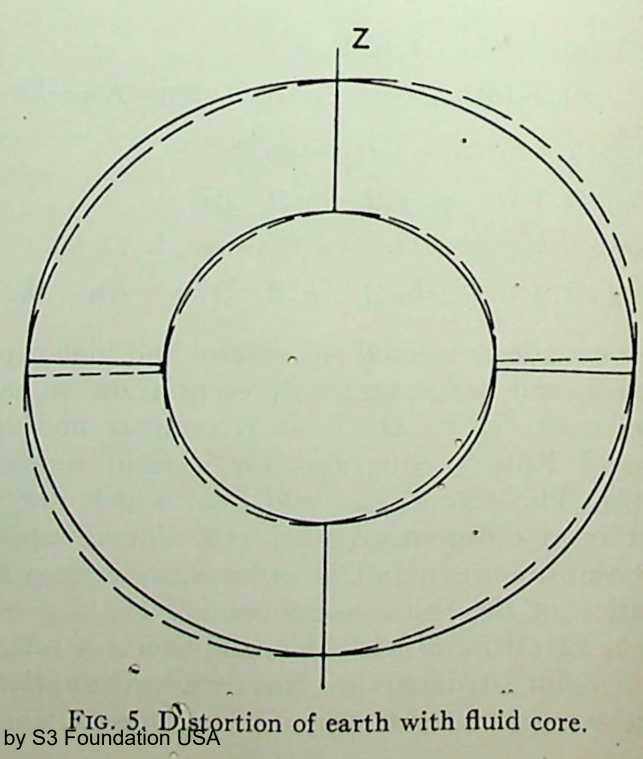


FIG. 5. Distortion of earth with fluid core.

to the simple shear of Fig. 4. The distortion of Fig. 5 accomplishes the same sort of displacement of the equator relative to the poles (in first order in the bending displacements because of the original curvature of the surface) and the bending involves shear over surfaces and through depths of the same order of magnitude as are involved in Fig. 4, so a somewhat similar result would be expected from a proper treatment of the more complicated actual distortion. It seems likely that an equivalent distortion is considerably easier in Fig. 5 than in Fig. 4, so that greater rigidity would be required, which would come closer to agreement with the empirical internal rigidity of the earth, as noted in the above discussion of the Klein-Sommerfeld treatment.

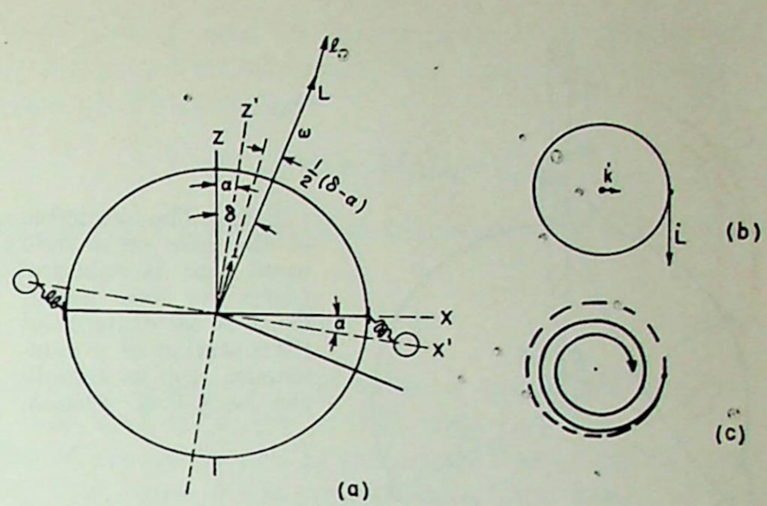


FIG. 6. A model to illustrate the viscous damping of the earth's nutation in terms of the slipping of an equatorial band in the xy plane to which is attached the ring representing the equatorial bulge.

FURTHER DISCUSSION OF THE COUPLED GYROSCOPES

The problem of the coupled gyroscopes has much in common with the problem of a pair of simple coupled oscillators. For fairly weakly coupled oscillators, the lower of the two normal frequencies is only a little lower than the lower of the original frequencies (defined for one body by holding the other at rest). If the other original frequency is considerably higher, the other body moves in this normal mode with relatively very small amplitude, so that the weak coupling spring by being stretched with relatively large amplitude can greatly affect its motion.[4] With strong coupling there is a similar behavior but the disparity of amplitudes is not so great. The solution we have discussed, (14) with $\alpha f \ll D$, corresponds to a mode in which the amplitude of the elastic yielding in the earth is rather small and the speed of nutation not much affected thereby.

In the solution (14), the angular displacement vector α is normal to the normal component of the angular velocity $\mathbf{\Omega}$. The linear displacement caused by $\boldsymbol{\alpha}$ is normal to $\boldsymbol{\alpha}$, so the maximum displacement of the ring away from the equatorial plane is in the azimuth of $\Omega_\perp$. This corresponds to the Klein-Sommerfeld assumption that the elastic reaction is in equilibrium with the instantaneous position of the rotation axis.

In seeking a singly-periodic solution (14) with K large and α small, we have excluded the other normal mode in which α would oscillate with an amplitude larger than δ and with a considerably higher frequency dictated more by the vibration than by the rotation. Such a vibration would be expected to be highly damped and not appear as a component of the earth's motion.

SIMPLEST EXPLANATION OF THE SLOWED NUTATION

For the sake of giving a further simplified exposition of the slowing-down of the nutation by the elastic deformability, let us consider a modification of the model in Fig. 3 in which the central body is a sphere, as in Fig. 6(a), so that the ring represents the excess matter in the equatorial bulge. The motion of the ring represents the motion of much more matter than that in the equatorial bulge, as we have seen, so the amplitude of α may be unrealistically large, but the model nevertheless reproduces the main features of the effect. The equatorial plane xy is defined as the plane at which the springs are attached to the sphere. The angular momentum $\mathbf{L}$ of the sphere is along the axis of rotation, $\boldsymbol{\omega}$. The principal axes x' and z' of the ring are inclined at an angle α to x and z because of the centrifugal force that tends to set the ring in a plane normal to $\boldsymbol{\omega}$ and this stretches the springs. The angular momentum $\mathbf{l}$ of the thin ring lies halfway between $\boldsymbol{\omega}$ and z', since $I_{x'x'}=\frac{1}{2}I_{z'z'}$, that is, $\mathbf{l}$ makes an angle $\frac{1}{2}(\delta-\alpha)$ with L, where δ is the angle between the axis of rotation $\boldsymbol{\omega}$ and the polar axis z. The vector $\mathbf{L}$ is moving out of the plane of Fig. 6 at a rate $d\mathbf{L}/dt=-(d\mathbf{l}/dt)$ because of the torque applied by the spring, which may be thought of also as resulting from the centrifugal force of the matter in the ring which is above the plane normal to ω (by an amount measured by the angle $\delta-\alpha$) on the right side of Fig. 6(a), and below it on the left side. The magnitude of this rate of change of angular momentum is thus

$$|d\mathbf{L}/dt| = |d\mathbf{l}/dt| = |\mathbf{l}\times\boldsymbol{\omega}| = \tfrac{1}{2}(\delta-\alpha)l\omega. \qquad (22)$$

This is the rate of change of $\mathbf{L}$ in the body-fixed coordinate system as well as in the space-fixed system because $\mathbf{L}$ is along the instantaneous axis of rotation. It is normal to the plane containing z and $\mathbf{L}$, so in the body-fixed system $\mathbf{L}$ circles around the z axis at this rate proportional to $(\delta-\alpha)$. Thus the elastic yielding which produces the angle α reduces the rate of nutation.

One may instead think of the torque causing $d\mathbf{L}/dt$ in terms of the centrifugal force of the parts of the ring shown in Fig. 6(a), which operates with a smaller "lever arm" than if x' were coincident with x.

VISCOUS DAMPING OF THE NUTATION

Let us now modify the model by assuming that the springs are attached not directly to the sphere but to a

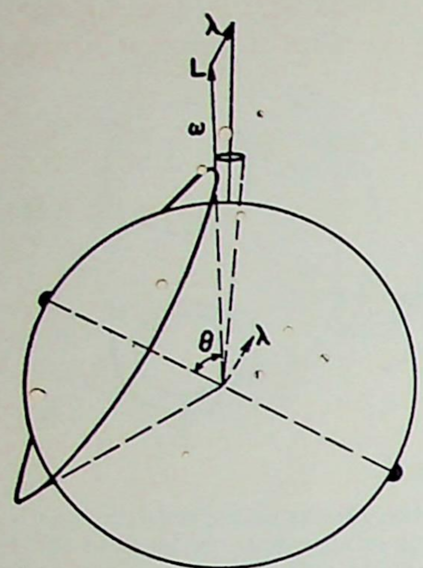


FIG. 7. The wandering of the pole on a rigid earth that is spherical except for local mountains is an example of the nutation of a symmetrical top as seen in the body-fixed system.

snug light band around the equator which is lubricated by a layer of viscous grease that permits slow slipping. This provides a plausible representation of the viscous flow in the earth. If the system is rotated for a long time about a fixed axis, this axis is gradually approached by the axis of symmetry, as is expected also of the viscous earth.

The unit vector $\mathbf{k}$ along the z axis is defined as being normal to the plane of the slipping equatorial band. It is thus not fixed relative to the sphere, but in a body-fixed system has a rate of change $\partial\mathbf{k}/\partial t$ proportional to the torque transmitted by the viscous connection from the sphere to the ring, that is

$$|\partial\mathbf{k}/\partial t| = K|d\mathbf{l}/dt| \tag{23}$$

$\partial\mathbf{k}/\partial t$ being in the plane of Fig. 6(a) and $d\mathbf{l}/dt$ normal to it. The vector $\partial\mathbf{k}/\partial t$ represents a motion of the symmetry axis toward the axis of rotation, so that $|\partial\mathbf{k}/\partial t| = -\dot{\delta}$, and $d\mathbf{l}/dt$ is given by (22) with α small and proportional to δ. Thus we have

$$\dot{\delta} = -K_1\delta, \tag{24}$$

with $K_1 \approx \frac{1}{2}Kl\omega$. There is then an exponential decay of δ and the instantaneous axis of rotation ω follows a spiral path on the surface of the sphere as suggested in Fig. 6(c).

PART II. WANDERING OF THE EARTH'S POLAR AXIS

The theory of polar wandering advanced by Gold[1] and formulated by Burgers[5] postulates mountain building or other asymmetric contribution to the moments and products of inertia of the earth as the motive power of the wandering. Let us consider the simple and unrealistic case of a rigid spherical earth with two "mountains" at opposite ends of a diameter at an angle θ from the polar axis, thus keeping the center of mass at the center for simplicity, as in Fig. 7.

We again let $\mathbf{L}$ be the angular momentum of the sphere and now let $\boldsymbol{\lambda}$ be the angular momentum of the mountains, which lies in that plane and normal to the line between the mountains, as in Fig. 7. Its magnitude is

$$\lambda = mR^2\omega \sin\theta \cos\theta, \tag{25}$$

where m is the total mass of the mountains. From $|d\mathbf{L}/dt| = |\boldsymbol{\lambda}\times\boldsymbol{\omega}|$, as in Eq. (22), we have

$$|d\mathbf{L}/dt| = mR^2\omega^2 \sin\theta \cos^2\theta, \tag{26}$$

the vector being normal to the plane containing $\boldsymbol{\omega}$ and the preferred body-fixed axis, the line between the mountains, leading again to the rolling-cone construction, this time as shown in Fig. 7. Equation (26) gives the rate at which the pole wanders about the body-fixed cone with half-angle θ about this axis. The wandering becomes relatively slow as θ approaches 90° and the path of wandering approaches a great circle. The narrow cone in Fig. 7, with axis $\mathbf{L}+\boldsymbol{\lambda}$, is of course fixed in space and the wide cone rolls on it. This is just a case of the familiar free nutation of a symmetrical top, but is here described in this unconventional way for future reference.

MOUNTAIN BUILDING ON A VISCOUS SPHEROIDAL EARTH

On an earth that is spherical except for some local mountain building, the pole may thus wander with great ease. The difficulty about finding a satisfactory theory of polar wandering lies in the fact that it is impeded by the quite appreciably nonspherical figure of the earth. The equatorial bulge is so massive (compared with a mountain range, for example) that it is difficult to perturb the motion from the Euler-Chandler nutation which leaves the rotation axis near the geometrical pole, unless the viscosity is so low as to permit easy readjustment of the figure.

Let us consider a model in which the "mountains" of Fig. 7 are added to the sphere having the slip-ring structure of Fig. 6(a) in place of the equatorial bulge, as in Fig. 8(a). It is not essential that we retain the elasticity of the spring, which makes only a moderate difference in the period of the nutation, as we have seen. We may for simplicity consider it rigid. It is

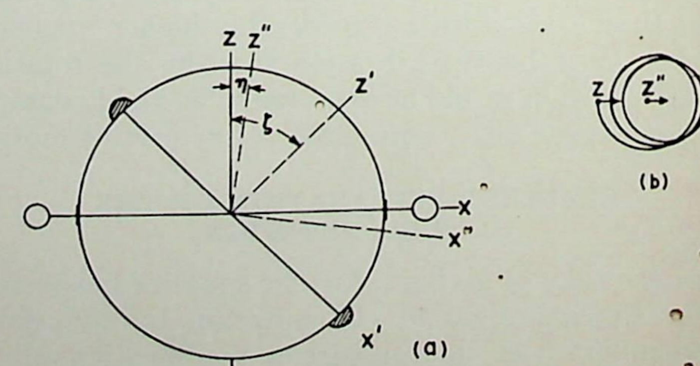


FIG. 8. Model to illustrate wandering of pole caused by mountainous or continental asymmetry and opposed by an equatorial bulge subject to viscous readjustment.

essential that we permit the equatorial band supporting the ring to slip with a viscous drag on the surface of the sphere, for it is this viscous readjustment that permits polar wandering.

But first we consider the motion without slipping. We then have a slightly asymmetrical top. If the mountains were added at the equator, we would have a slightly asymmetrical top with the same principal axes as before. The nutation would then be almost the same as for the symmetrical top, but with slightly elliptical rather than exactly circular rolling cones. [This may be seen, for example, by noting that when the mountains are in the plane of Fig. 6(a), they help the ring by centrifugal force to produce $d\mathbf{L}/dt$, whereas when they are normal to this position they do not.] If the mountains are added away from the equator or poles, they have also the more significant effect of altering the principal axes slightly. Let the diameter through the mountains be the x' axis and let the z' axis normal to it make an angle ζ with the z axis. Then a new pair of axes x'', z'' rotated in the plane by an angle η from x, z has the product of inertia

$$I_{x''z''}=(I_{xx}-I_{zz})\cos\eta\sin\eta + mR^2\cos(\zeta-\eta)\sin(\zeta-\eta)=0 \quad (27)$$

which must vanish if these be the new principal axes. In the approximation $mR^2 \ll I_{zz}-I_{xx}$ and consequently small η, we have

$$\eta = mR^2\cos\zeta\sin\zeta/(I_{zz}-I_{xx}). \quad (28)$$

Without slipping, the motion of the axis will thus be a nearly circular nutation about an axis displaced from the geometrical axis z by a small angle η given by (28). If for example the original rotation were one about the z axis, the subsequent motion would be a nutation about an approximately circular cone of half-angle η passing through the z axis at one edge. The displacement δ of the axis of rotation from the z axis may be described (approximately) as a constant angular displacement η plus a uniformly rotating angular displacement.

Now with slow viscous slipping of the equatorial band permitted, the rate of change of the z axis may be presumed linearly proportional to δ and the uniformly and fairly rapidly rotating component of δ has no cumulative effect. It merely damps the rotating part itself. The cumulative rate of wandering of the z axis toward the z' axis (which means away from the mountains) is thus proportional to η as given by (28), and as ζ varies only slowly so does η. The average axis of rotation, or polar axis, z'', thus also wanders slowly keeping just ahead of the z axis, as suggested in Fig. 8(b) (which is drawn for the case in which the initial rotation is about the z axis, though the rotary part may instead be initially much larger). Thus the cumulative part of the slipping takes place as if the axis of rotation were z'', in which case the angular momentum of the ring, $\mathbf{l}$, makes an angle $\frac{1}{2}\eta$ with z'', giving us $|d\mathbf{l}/dt| = \frac{1}{2}\eta\omega l$. This is equal to the torque that is causing the ring to slip and thus causing the z axis to move toward the z' axis. Thus we have

$$\dot{\zeta}=K|d\mathbf{l}/dt|=\tfrac{1}{2}K\omega l\eta=K_1\eta \quad (29)$$

with the same constant K as appeared similarly in Eq. (23), or

$$\dot{\zeta}=-K_2\zeta \quad (30)$$

with $K_2=\frac{1}{2}K\omega l(\eta/\zeta)$. The ratio of angles η/ζ is given by (28) and in it for a thin ring $I_{zz}-I_{xx}=c-a=a=\frac{1}{2}m_rR^2$ as mentioned in connection with Eq. (8), m_r being the mass of the ring or equatorial bulge. If we put, roughly, $\cos\zeta\sin\zeta/\zeta\approx\frac{1}{2}$, we have

$$K_2\approx\tfrac{1}{2}K\omega lm/m_r=(m/m_r)K_1. \quad (31)$$

Here K_1 gives the rate of decay of the nutation, as in (24). Thus the relative decay rate of ζ is smaller than the decay rate of δ by about the ratio of the mass of the "mountains" to the mass of the equatorial bulge of the earth.

Gold[6] mentions the possibility of continential rising by amounts between 30 meters in 10^5 and 3 meters in 10^6 years. The actual rate of rise of continents is probably largely due to isosastatic readjustment to the removal of overburden by erosion or glacial melting. In the case of glacial melting, the material removed is distributed over the seas and the main effect is a replacement of the locally removed material by the rising crust, leaving only a small differential effect once equilibrium is reached. In the case of erosion, the removed material is deposited around the edges of the continent where it should cause a sinking to compensate its arrival, in the same way that the rise within the continent compensates the departure of the material, again leaving only a small differential effect. (Because of this approximate isostatic compensation, it does not seem plausible to make the assumption of a continuously rising "mountain" such as led Munk[6] to propose an angle of accelerated polar wandering proportional to the square of the time.)

If we consider the simple primary effect of a rise of an area as our "mountain," it seems safer to assume the smaller figure 3 meters rise over a continent of an area of say, 2 10^7 square kilometers, giving $m/m_r\approx 2\ 10^{-5}$, but even this is plenty to produce a fairly rapid polar wandering. With (30), it gives K_1/K_2 times the value 4 years for the decay of δ mentioned above,[7] or 2 10^5 years for wandering by the order of ζ, which may be of the order of one radian. This is much shorter than the time since the Eocene during which the pole is thought to have wandered from the neighborhood of Hawaii to its present position, so a more modest uplift than this if sustained so long would suffice, or a considerably slower plastic flow than we have assumed from the apparent damping of the nutation. Temporary,

glacial loadings may be a hundred times this great so as to require only 2000 years for polar wandering through large angles, and such a local loading may remain out of isostatic equilibrium for such a period, making it possible to accumulate the long-term wandering as a series of quick steps associated with successive glacial ages.

If we instead assume isostatic equilibrium and treat the differential effect, we may, for example, plausibly take a continental area to consist of rock 20 km deep having a mean density 10% less than that under the oceans, so the continent would float about 2 km higher than the ocean floor. The center of mass of this rock is higher by $dR=1$ km than would be that of the denser rock it has displaced. Its contribution to the quantity $\int \rho r^2 d(\text{vol})$ at this angle, which is the significant quantity in determining the moments and products of inertia, is greater than that of the matter displaced by $2(\text{mass})RdR$ which is to be compared with mR^2 when we simply add a mountain of mass m. Thus the differential effect is equivalent to adding a layer of thickness $(2dR/R)20$ km $=7$ meters, about twice as great as the "mountain" of continental area just considered. We see, then, that polar wandering at something like a radian in 10^5 years is to be expected from the very existence of continents, unless the pole has become "trapped," as Gold expresses it, in a position such that the various continents or the various inhomogeneities of the earth's crust just balance one another and leave the principal axis practically coincident with the axis of rotation.

Gold points out that such a "trap" could be stabilized by the melting of ice near the poles which is influenced by the position of the pole and in turn influences the position of the pole through the mechanism here discussed. That the sign of the feedback is correct for stability is seen from the fact that in Fig. 8 the axis of rotation z'' moves away from the mountains, permitting the mountains or other excess mass in one quadrant to move toward the equator. (Incidentally, this tends to maximize the moment of inertia I about the axis of rotation, which makes the kinetic energy $J^2/2I$ a minimum as a result of energy dissipation through viscous flow.) As the pole drifts away from the quadrant in which there is excess mass, the part of that quadrant near the pole becomes warmer and ice melts there, reducing the excess mass, while freezing and ice accumulation takes place on the opposite side of the pole where there was a deficit of mass. This self-correcting mechanism is operative either in a region like the Arctic where the pole is near the middle of a sea surrounded by continental areas on which ice may accumulate, or in a region like the Antartic where the pole is near the middle of an ice-bearing land mass. It is very plausibly the simultaneous occurrence of both that accounts for the stability of the pole near its present position.

MARINE-CURRENT THEORY OF POLAR WANDERING

Before Gold's paper had suggested the mechanism discussed above, there seemed to be another possible explanation of the earth's polar wandering that grew out of some remarks of Enrico Fermi. During the autumn of 1953, at the informal and far-ranging seminar in the Institute for Nuclear Studies that now bears his name, he gave an estimate to show that the angular momentum of a large marine current such as the Gulf Stream is sufficient to influence the rotation of a spherical earth and cause appreciable polar wandering. He later became aware of the essential difficulty presented by the very considerable ellipticity of the earth and at the time of our last discussion some months later he seriously doubted that the viscous flow could be fast enough to permit wandering through a large angle in geologic times. If we accept the viscous flow suggested by the damping of the nutation (or even ten times less) it turns out, as we shall see, that the marine currents do provide an ample source of the slow polar wandering but are much less important than the asymmetry mechanism of Gold.

As figures to give the order of magnitude of such a marine-current angular momentum we may take a flow of 100 cubic miles per hour, or 10^8 meters^3/sec, of water around a path of mean radius 2000 km, giving an angular momentum 3×10^{24} kg meters^2/sec. This is the angular momentum of a rather large mass of water, say 400 miles wide and $\frac{1}{4}$ mile deep, traveling at the rather slow speed of 1 mile per hour around a radius about a third of the earth's radius.[10]

Now let us go back to Fig. 7. There we had an angular momentum λ, due to the mountains moving with the earth's rotation, which reacted on the spherical earth to cause polar wandering. With a marine current instead of mountains to perturb the symmetric top, the vector λ in that diagram may now be taken to represent the angular momentum of the marine current which has the same immediate effect, and the axis of rotation again starts to move on a path normal to the plane of λ and ω. The difference is that the preferred body-fixed axis is now the direction of λ, since the marine current is constrained to move about the same body-fixed path, and the path of ω is a large body-fixed cone now having as its axis the direction of the angular momentum λ, rather than the line between the mountains.

A "mountain" consisting of a layer of rock of specific gravity 4 covering a continent to a thickness of three meters, as we have assumed above, has a mass

[10] Large air currents have angular momenta of this same order of magnitude, since they have masses about 10^2 smaller (densities about 10^3 smaller but larger volume) and velocities 10^2 larger. While affected by continental masses. they are not confined thereby as marine currents are, and it seems likely that the much more rapid and unconfined shifting of air currents leads to a cancellation that makes them relatively ineffective in shifting the earth's axis. In any case, we may estimate the combined effect in terms of the marine currents alone.

$m = 2\times 10^{17}$ kg which in Eq. (25) with $\sin\theta \cos\theta = \frac{1}{2}$ gives an angular momentum $\lambda = 4\times 10^{26}$ kg meters^2/sec. We have seen the differential effect of a floating continent in isostatic equilibrium is about this great (our estimate was twice this) whereas the effect of a marine current is only about a hundredth of this. In this comparison, the marine current has a mass 3×10^{18} kg, more than ten times that of the continental "mountain" which, however, moves with the earth's rotation a thousand times as fast as the marine current flows relative to the earth (it being, of course, only its *additional* angular momentum due to its flow that contributes because the water involved is otherwise just part of the symmetrical earth).

If we are simply looking for a mechanism capable of causing polar wandering through a large angle in about 10^8 years, as seems to be required by geological evidence, we have found the asymmetry mechanism to be more than adequate with about a factor 10^3 to spare, so the marine-current mechanism alone, although a hundred times weaker, is still ample. It is, however, overwhelmed by the effect of possible asymmetries in the crust and expected to play only a relatively small role in determining the final balance. There are, of course, several marine currents (just as there are several continents) and the vector sum of their angular momenta, which is what matters, is probably not much greater than that of one marine current. The circulation of the atmosphere, which is less massive but more rapid, has a comparable small effect.[10]

Presentation of the nutation and wandering has not here been complicated by inclusion of the effects of the fluid core. Treating it as a further gyroscope coupled viscously to the mantle in such a way as to follow in the nutation rather than drive it, Bondi and Gold[11] have shown that its moment of inertia is not great enough for it to be responsible for the damping.

[11] H. Bondi and T. Gold, Monthly Notices Roy. Astron. Soc. **115**, 41 (1955).

Broadening and Shift of Spectral Lines Due to the Presence of Foreign Gases*

SHANG-YI CH'EN† AND MAKOTO TAKEO‡

Department of Physics, University of Oregon, Eugene, Oregon

TABLE OF CONTENTS

I. INTRODUCTION

THE modification of the spectral lines of an emitting or absorbing particle due to collisions with other particles as exhibited by broadening, shift, asymmetry, and the occasional appearance of certain diffuse bands, has opened up a new way to study the perturbations of energy levels. The variation of interatomic and intermolecular forces as a function of interatomic distance as determined from these effects has been increasingly attracting the attention not only of theoretical and experimental physicists but also of those who are working on certain basic problems of chemistry, genetics and astrophysics.

The first review article on the causes of broadening of spectral lines was written by Weisskopf[1] in 1933. Margenau and Watson[2] made an extensive review of the theories and experimental observations on the pressure effects on spectral lines developed up to 1936. Schulz,[3] Budini,[4] and Unsöld[5] wrote corresponding reviews in later years. There has been a very considerable amount of work done within the past thirteen years, both theoretical and experimental, and the need of another review is quite obvious.

* Supported by the National Science Foundation, NSF-G1872.
† In memory of his father, T. W. Ch'en, July, 1884–July 20, 1956.
‡ Formerly at Department of Physics, Defense Academy, Yokosuka, Japan; now at Physics Department, Portland State College, Portland, Oregon.

[1] V. Weisskopf, Physik. Z. 34, 1 (1933).
[2] H. Margenau and W. W. Watson, Revs. Modern Phys. 8, 22 (1936).
[3] P. Schulz, Physik. Z. 39, 412 (1938).
[4] P. Budini, Nuovo cimento 16, 86 (1939).
[5] A. Unsöld, Vjschr. Astron. Ges. 78, 213 (1943).

This review article is intended to constitute a sequel to Margenau and Watson's article, to bring it up to 1956. To restrict the length of this paper, only pressure effects due to neutral particles are discussed. The effects of electron and ion collisions, high-temperature broadening, and, consequently, observations with shock waves are postponed for consideration in a separate review. The modification of both atomic and molecular spectral lines in the visible, ultraviolet, infrared, and microwave regions is discussed. In Chapter III, the materials of Secs. B and C are purposely given separately from Sec. A for emphasis. During the course of the preparation of this article, Sobel'man and Robin and Robin[6] gave a review of the theories and a bibliography of the pressure broadening of atomic lines respectively. The present article is an entirely independent review covering also molecular lines and includes a review of experimental observations.

II. THEORY

Since Lorentz presented his theoretical treatment[7] of the problem of the broadening of spectral lines in 1906, a number of different approaches have appeared. The relations between them are sometimes difficult to see and it is not always clear whether they describe the same effect. It may be, therefore, of interest to review these treatments. However, those appearing in the review article published by Margenau and Watson[2] are mainly excluded.

A. CAUSES PRODUCING BROADENING AND SHIFT OF SPECTRAL LINES

Every influence on a radiating or absorbing atom (or molecule) affects its spectral lines in one way or another. The effect of its own thermal motion, which broadens the lines statistically, is known as the Doppler effect.[8,9] In the case of emission or resonance absorption the process of radiation damping,[9] or a finite width of each of the two energy levels associated with spectral lines, is responsible for the natural width.[10] If the atom (or molecule) is in a gas of the same or of different kind, the spectral lines are broadened and sometimes are shifted with asymmetry of the line shape. The line broadening due to the first two effects is independent of pressure, while that due to the last effect (pressure effect) depends on the density of the gas and is, under ordinary experimental conditions where the pressure is more than a few cm of Hg, much larger than the first two effects in the visible down to the microwave region. Therefore, the Doppler and radiation damping effects on line shapes will be entirely neglected in this discussion,[11] although for emission spectra in flames the Doppler effect should be added.[12]

The radiating or absorbing atom (or molecule) in a gas is exposed to the interaction with surrounding atoms (or molecules) which are moving under thermal motion. The essential cause of line broadening lies in the finite difference of interaction energies of the radiating atom (or molecule), in the initial and final states involved in the radiation process, with a colliding atom (or molecule). It is usually not easy to find the interaction energy over the whole range of the distance, including excited states. Thus, it is sometimes necessary to use the approximate Lennard-Jones potential energy,

$$V = -\alpha R^{-6} + \beta R^{-12},$$

which has been found to be useful in the kinetic theory of gases. When a more detailed discussion is required, the potential energy of the classical electrostatic interaction between atoms or molecules is expanded in a series in R^{-1}. This is equivalent to the procedure of expanding the two charge distributions in terms of the electric multipole moments in the expression of the electrostatic interaction. However, in this expansion care must be taken to determine whether or not the two charges overlap. Otherwise, the resulting expansion in series may diverge for all R.[13]

Referring to the coordinate systems, x_i, y_i, z_i, and x_j, y_j, z_j parallel to each other with z_i and z_j in the direction of R and with the origin at the nuclei of each charge distribution i or j, $V(R)$ is given by[14]

$$\begin{aligned} V(R) \sim & -\frac{1}{R^3}\sum_{ij} e_i e_j [2z_i z_j - x_i x_j - y_i y_j] \\ & + \frac{3}{2}\frac{1}{R^4}\sum_{ij} e_i e_j [r_i^2 z_j - z_i r_j^2 \\ & + (2x_i x_j + 2y_i y_j - 3z_i z_j)(z_i - z_j)] \\ & + \frac{3}{4}\frac{1}{R^5}\sum_{ij} e_i e_j [r_i^2 r_j^2 - 5z_i^2 r_j^2 - 5r_i^2 z_j^2 \\ & - 15 z_i^2 z_j^2 + 2(4z_i z_j + x_i x_j + y_i y_j)^2] + \cdots \end{aligned} \quad (1)$$

for electrically neutral and tightly bound charge distributions at a large distance $R (R > r_i + r_j)$. If the two charges overlap, the integral over the charge distribution should be cut off at the distance where the two charges start to overlap, in order to obtain a convergent potential for all R.

---

[6] L. I. Sobel'man, Uspekhi Fiz. Nauk 54, 551 (1954); S. Robin and J. Robin, J. phys. radium 17, 143 (1956).
[7] H. A. Lorentz, Proc. Acad. Sci. Amsterdam 8, 591 (1906).
[8] L. G. Henyey, Proc. Natl. Acad. Sci. 26, 50 (1940).
[9] D. G. Kendall, Z. Astrophys. 16, 308 (1938).
[10] V. Weisskopf and E. W. Wigner, Z. Physik 63, 54 (1930); 65, 18 (1931); F. Hoyt, Phys. Rev. 36, 860 (1931); L. Spitzer, Jr., Phys. Rev. 55, 361 (1939); Physica 7, 133 (1940); E. Arnous and W. Heitler, Proc. Roy. Soc. (London) A220, 290 (1953).
[11] However, in the microwave region at a low pressure the Doppler effect will bother. Reduction of this effect in observations has been studied by M. W. P. Strandberg and H. Dreicer, Phys. Rev. 94, 1393 (1954).
[12] S. S. Penner and R. W. Kavanagh, J. Opt. Soc. Am. 43, 385 (1953); L. Huldt and E. Knall, Z. Naturforsch. 9a, 663 (1954).
[13] F. C. Brooks, Phys. Rev. 86, 92 (1952).
[14] H. Margenau, Phys. Rev. 38, 747 (1931).

The recently developed expansion[15] of $V(R)$ has made it possible to systematize and simplify the study of the interaction of complex charge distributions even when they overlap. The expansion of $V(R)$ is in terms of spherical harmonics $Y_\lambda^\mu(\theta,\phi)$ corresponding to multipole moments of the charge distributions. If one refers to a spherical coordinate system with origin at the center of mass and the polar coordinate in the direction of R, the coordinates of charges e_i, belonging to one atom or molecule, are (r_i,θ_i,ϕ_i) and those of charges e_j, belonging to the other, are (r_j,θ_j,ϕ_j). For no overlapping of charge distributions, i.e., $R>r_i+r_j$, Carlson and Rushbrooke[15] gave the following expansion:

$$V(R)=\sum_{ij} e_ie_j \sum_{\lambda=0}^{\infty}\sum_{\lambda'=0}^{\infty} \frac{r_i^{\lambda}r_j^{\lambda'}}{R^{\lambda+\lambda'+1}} \frac{(-1)^{\lambda_L}4\pi(\lambda+\lambda')!}{[(2\lambda+1)(2\lambda'+1)]^{\frac{1}{2}}}$$

$$\times \sum_{\mu=-\lambda_L}^{\mu=+\lambda_L} \frac{y_\lambda^\mu(\theta_i,\phi_i)y_{\lambda'}^{-\mu}(\theta_j,\phi_j)}{[(\lambda+\mu)!(\lambda-\mu)!(\lambda'+\mu)!(\lambda'-\mu)!]^{\frac{1}{2}}}. \quad (1')$$

Here λ_L is the lesser of λ and λ'. Buehler and Hirschfelder[15] have obtained the rather complicated expression for the case when the two charges overlap each other. Since this type of mathematical approach includes every case of charge distributions in a convergent form, it is of great value in the evaluation of Coulombic integrals in the quantum-mechanical treatment of short-range intermolecular forces.

The potential (1) or (1′) is considered as a perturbation upon the system and perturbation theory or the variational method may be employed to evaluate the resultant shift in energy levels, which is identified with the potential energy of the intermolecular force. If ions are not considered,[16] V starts with $1/R^3$ as in Eq. (1) in the approximation[17] of separate electronic, nuclear, and molecular motions. At low pressures (average R large), the predominant part of the interaction arises from the term of lowest power, and other higher terms may be neglected. On the other hand, at close distances interaction energies are quite complicated and will not be discussed here, although they are quite important.

In the discussion of long-range van der Waals' forces it is customary to consider four types of forces: (i) electrostatic forces such as those between electric dipoles, (ii) induction forces such as those between a polarizable nonpolar molecule, and an electric dipole, (iii) dispersion forces such as those between nonpolar molecules, and (iv) resonance forces realized by exchanges of excitation energy between molecules of the same kind. In the following paragraphs, however, the types of forces discussed in the several cases are only the more important ones.

1. Interaction between Atoms

a. Resonance Interaction between Identical Atoms

This is characteristic of the case when the interacting atoms, one of which is excited, are identical. Since they are identical, they can exchange the excitation energy (light quantum) with one another.[2,18] This resonance introduces twofold degeneracy (symmetric§ and antisymmetric) and a term,

$$V_{\text{res}}=\gamma(e^2\hbar f_{0n}/4\pi m_e\omega_0)\frac{1}{R^3}, \quad (2)$$

in the interaction energy. The numerical factor γ takes the value -2 if the magnetic quantum number m of the excited atom is 0, 1 if $m=\pm1$. f_{0n} is the oscillator strength corresponding to the transition from the normal to the excited state n, ω_0 the angular frequency of the spectra line concerned, and m_e the electronic mass. This term is not additive,[2,19,20] which makes it difficult to calculate in the case of high pressure where multiple interactions are involved, and has not been treated rigorously, although various[21] attempts have been made. However, according to the sum rule, the average of the interaction energy over all states of different m's is zero. The root mean square value over magnetic quantum numbers m in the J representation for a binary case was given by Foley[22] to show its J dependence, but Eq. (2) will be in error for the case when multiplet separation is comparable or less than the perturbation energy.

It is easy to see how the resonance effect affects the radiation process.[23] When an excited atom passes by an unexcited atom of the same kind, there must be a finite probability that the energy of excitation will be transferred from the first to the second atom, without the intervention of radiation. This results in a reduction

---

[15] C. Carlson and G. S. Rushbrooke, Proc. Cambridge Phil. Soc. **46**, 626 (1950); R. J. Buehler and J. O. Hirschfelder, Phys. Rev. **83**, 628 (1951); **85**, 149 (1952).

[16] Stark effect broadening of spectral lines due to ions, and dipole and quadrupole gases are considered by H. Holtsmark, reference 69; M. Born, *Optik* (Verlag Julius Springer, Berlin, 1933), p. 444; R. Rompe and P. Schultz, Z. Physik **119**, 269 (1941); S. L. Mandelshtam and N. N. Sololev, J. Exptl. Theoret. Phys. **20**, 323 (1950); I. I. Sobel'man and L. A. Vainshtein, Doklady, Akad. Nauk S.S.S.R. **90**, No. 5, 757 (1953); and by E. F. M. v.d. Held, Z. Physik **70**, 508 (1931). Also see, for instance, reference 94, and R. G. Breene, Jr., Revs. Modern Phys. **29**, 96 (1957), Chap. II.

[17] Hirschfelder, Curtiss, and Bird, *Molecular Theory of Gases and Liquids* (John Wiley and Sons, Inc., New York, 1954), p. 920.

[18] G. W. King and J. H. Van Vleck, Phys. Rev. **55**, 1165 (1939).

§ Only the symmetric states with respect to the exchange of the atoms can combine optically with ground states.

[19] For information on the nonadditivity in first-order interactions for three helium atoms, see P. Rosen, J. Chem. Phys. **21**, 1007 (1953).

[20] H. Margenau, Revs. Modern Phys. **11**, 1 (1939).

[21] J. Holtsmark, Z. Physik **34**, 722 (1925); J. Frenkel, Z. Physik **59**, 198 (1930); L. Mensing, Z. Physik **61**, 655 (1930).

[22] H. M. Foley and D. M. Dennison, Phys. Rev. **61**, 386 (1942); H. M. Foley, Phys. Rev. **69**, 616 (1946).

[23] W. Furssow and A. Wlassow, Physik. Z. Sowjetunion **10**, 379 (1936); W. V. Houston, Phys. Rev. **54**, 884 (1938). Also see R. G. Breene, Jr., Revs. Modern Phys. **29**, 122 (1957), Chap. V.

of the lifetime of the excited atom and a corresponding broadening of the line emitted.

b. Dispersion Interaction

The charge distribution of atoms in the ground states is usually nondegenerate with respect to parity, without electric dipole moments.[24] Thus, the first-order dipole-dipole interaction vanishes as does the term $1/R^3$ in Eq. (1). This is true in the system of identical atoms in the ground states or of atoms of different kind. The second order perturbation theory gives the asymptotic potential energy[20] between two atoms in the states k and l as,

$$V_{\text{dis}} = -\frac{1}{R^6}\frac{3}{2m_e^2}(\hbar e)^4 \sum_{\substack{k'l'\\(k\neq k',l\neq l')}} \times \frac{f_{kk'}f_{ll'}}{(E_{k'}-E_k)(F_{l'}-F_l)(E_{k'}+F_{l'}-E_k-E_l)} + 0\left(\frac{1}{R^8}\right), \quad (3)$$

where $f_{kk'}$ is the oscillator strength corresponding to the transition kk' for the first atom, $f_{ll'}$ for the second. The E's and F's are the energies of the various states for the two atoms, respectively. It must be noted that the calculation was done by neglecting the electron-exchange effect, by assuming that atoms are nondegenerate, that they are quite far away from one another, and that their wave functions do not overlap. This last assumption clearly allows the additiveness[25] of the force. As will be seen easily, the value of Eq. (3) is usually negative, and numerically much larger for the excited state than for the ground state, except for the highly excited states. The defect of Eq. (3) is, however, that it requires a knowledge of every energy level of the atoms concerned, which is hard to obtain. The previous calculations were improved by Buckingham,[20,26] who applied the variational method to the ground states, made use of wave functions obtained by the method of self-consistent fields, and included the effect of electron exchange. This result provides a method for obtaining the dispersion energy from the observed polarizabilities α_1 and α_2 of the two atoms with N_1 and N_2 electrons, respectively:

$$V_{\text{dis}} = -\frac{3}{2}\frac{a_0^{\frac{1}{2}}e^2}{R^6}\frac{\alpha_1\alpha_2}{(\alpha_1/N_1)^{\frac{1}{2}}+(\alpha_2/N_2)^{\frac{1}{2}}}. \quad (4)$$

a_0 is the radius of the first Bohr orbit. Massey and Burshop calculated the van der Waals constants for alkali metal-rare gas atom impact by this method and found reasonable agreement with those obtained by the molecular ray technique.[27]

The above formulas are not valid for degenerate states, which are always involved in the optical transition, although Eq. (3) may be thought of as an average over the degeneracy for such levels. Really, it is not easy to obtain the van der Waals constant for excited states and to compare the line broadening theory with observations. Recently, the authors[28] have calculated the van der Waals energy for the doublet P state and found that $^2P_{\frac{3}{2}}$ levels split according to the magnetic quantum number due to the electric quadrupole moment of the atom in the level. The ratio of magnitudes of van der Waals energies for $^2P_{\frac{3}{2}}{}^{|m|=\frac{3}{2}}$, $^2P_{\frac{3}{2}}{}^{|m|=\frac{1}{2}}$, and $^2P_{\frac{1}{2}}{}^{|m|=\frac{1}{2}}$, interacting with a spherically symmetric atom in the ground state, is 4:5:6. They claim that this split of the $^2P_{\frac{3}{2}}$ level is responsible for one aspect of fine structure pressure broadening.

Under ordinary experimental conditions, the resonance interaction can be neglected in the line broadening due to foreign gases, because the number of foreign atoms or molecules surrounding the optically active atom is far greater than the number of neighbors of its own kind.

c. Energy State of Atoms at High Pressure

There have been some calculations on the ground state of atoms under high pressure using as a model an atom enclosed in a box of a certain size which corresponds uniquely to the pressure.[29]

The present authors[28] applied perturbation theory to this model for the case of an alkali atom compressed by rare gas atoms and obtained the energy corresponding to the isolated nth state.

$$E = W_n + (n|V|n) + \sum_j{}' (n|V|j)(j|V|n)/(W_n - W_j), \quad (5)$$

where W_n is the eigenvalue of the unperturbed nth state. $\sum_j{}'$ implies the exclusion of the term $j=n$ from the summation. The matrix element is

$$(n,l,m,s|V|n',l',m',s') = \begin{cases} \alpha \int_{r_0}^{\infty} R_{nl}^*(r)R_{n'l}(r)r^2dr, & \text{if } l=l',\ m=m',\ s=s', \\ 0, & \text{otherwise,} \end{cases}$$

[24] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), p. 158.

[25] Even though the wave functions do not overlap, the third-order energy is not additive. See B. M. Axilrod and E. Teller, J. Chem. Phys. **11**, 299 (1943); B. M. Axilrod, J. Chem. Phys. **17**, 1349 (1949); **19**, 719 (1951); **19**, 724 (1951).

[26] R. A. Buckingham, Proc. Roy. Soc. (London) **A160**, 113 (1937); Proc. Roy. Irish Acad. **45**, 31 (1938).

[27] N. S. W. Massey and R. H. S. Burshop, *Electronic and Ionic Impact Phenomena* (Clarendon Press, Oxford, 1952), p. 375.

[28] M. Takeo and S. Y. Ch'en, Phys. Rev. **93**, 420 (1954).

[29] Michels, deBoer, and Bijl, Physica **4**, 981 (1937); A. Sommerfeld and H. Welker, Ann. Physik (5) **32**, 56 (1938); S. R. deGroot and G. A. tenSeldam, Physica **12**, 669 (1946); **18**, 891, 905, and 909 (1952); T. L. Cottrell, Trans. Faradary Soc. **47**, 337 (1951).

where α stands for the dimension: $\alpha=1$ in atomic units of energy, decided upon by a comparison of the above calculation with the exact calculated value[29] for hydrogen. $R_{n1}(r)$ is the radial part of the wave function. r_0 is the radius of the sphere. The energy increases with decreasing r_0. The energy change is greater for the excited state than for the ground state at the same r_0. Since the pressure is given by $p=-dE/(4\pi r_0^2 dr)$, the corresponding hydrostatic pressure on the atom is easily calculated.

2. Interaction between Molecules

In order for an interaction to be effective for line broadening, the perturbation must be different for two levels in radiative transition. Any vibrational or rotational levels belonging to a specified electronic level have nearly the same second-order dispersion forces. Thus, in molecular spectra, the effect of dispersion forces can be neglected in many cases. Furthermore, it often happens that the separation of successive rotational levels is smaller than the thermal energy. For this reason, it must be noted that the ordinary second-order perturbation theory cannot be used in some cases.

Since molecules have more internal degrees of freedom than atoms, the calculation of multiple interactions is quite difficult. The following discussion is restricted to the binary interaction. Recently, however, Matossi[30] calculated the multiple interaction of identical polar molecules by considering the problem as the coupling of oscillators with essentially the same eigenfrequency ω_0 after the method of Holtsmark[21] for the atomic case. He showed that the multiple interaction gives rise to a splitting of the frequencies by an amount which depends on the distance and the orientation of the oscillating dipoles, thus resulting in distributed frequencies which can be considered as equivalent to line broadening. With a doubtful assumption that the frequencies are distributed according to the Gaussian error curve, he obtained a result that the half-width at half-intensity of the distributed frequencies is given by

$$\frac{1.18}{24\pi^{\frac{3}{2}}c^2}(8)^{\frac{1}{2}}\frac{\epsilon_e^2 N^{\frac{1}{2}}}{m r_0^{\frac{3}{2}}\omega_0}\ \mathrm{cm}^{-1},$$

where r_0 is the molecular diameter, c the velocity of light, m the reduced mass of the molecules, ϵ_e the effective charge defined as the ratio of dipole moment to atomic distance, and N the number density of the molecules. This $\sqrt{N}$ dependence of half-width was used to explain several experimental results,[31] especially those for the infrared absorption of water.[32]

[30] F. Matossi, Phys. Rev. **76**, 1845 (1948).
[31] B. Trumpy, Z. Physik **34**, 715 (1925); **40**, 594 (1927); Hosche, Polanyi, and Vogt, Z. Physik **41**, 587 (1927).
[32] W. M. Elsasser, Harvard Meteorol. Stud. No. 6 (1942), p. 46; F. Matossi and E. Rauscher, Z. Physik **125**, 418 (1949).

a. Interaction between Symmetric Top Molecules (NH_3, PH_3, CH_3Cl, CHF_3, CH_3F, PF_3, CF_3Cl, $CHBr_3$ etc.)

The intermolecular forces (electrostatic) between symmetric top molecules carrying a permanent electric dipole μ along their figure axis can be easily obtained by using the expansion (1). Assuming the two molecules in the rotational states J (total angular momentum), M (equatorial), K (symmetric top molecule quantum number), and J', M', K', respectively, we find the required matrix elements[33] are

$$(J,K,M|\cos\theta|J,K,M)=KM/[J(J+1)],$$

$$(J,K,M|\sin\theta e^{\pm i\phi}|J,K,M\pm1)$$
$$=\mp K/[J(J+1)]\cdot[(J\pm M+1)(J\mp M)]^{\frac{1}{2}}. \quad (6)$$

Then, the formula (1) leads, in its asymptotic form,[34] to

$$(J,K,M_j J',K',M_j'|V_{\mathrm{st}}|J,K,M_i,J',K',M_i')$$
$$=\frac{\mu\mu'}{R^3}\frac{KK'}{J(J+1)J'(J'+1)}$$
$$\times[-2M_iM_i'\delta(M_j,M_i)\delta(M_j',M_i')$$
$$+\tfrac{1}{2}\{(J-M_i+1)(J+M_i)(J'-M_i')$$
$$\times(J'+M_i'+1)\}^{\frac{1}{2}}\delta(M_j,M_i-1)\delta(M_j',M_i'+1)$$
$$+\tfrac{1}{2}\{(J+M_i+1)(J-M_i)(J'+M_i')$$
$$\times(J'-M_i'+1)\}^{\frac{1}{2}}\delta(M_j,M_i+1)\delta(M_j',M_i'-1)]. \quad (7)$$

Again, the sum rule shows that the average value over i and j vanishes. The mean absolute value may be approximated by its root-mean-square value,[34]

$$\langle V_{\mathrm{st}}^2\rangle^{\frac{1}{2}}(J,K;J',K')$$
$$=\left(\frac{2}{3}\right)^{\frac{1}{2}}\frac{KK'}{[J(J+1)J'(J'+1)]^{\frac{1}{2}}}\frac{\mu\mu'}{R^3}. \quad (8)$$

This can be used for two molecules of the same or different kind.

Since the selection rule for optical transition is $\Delta J=\pm1$, a molecule in the rotational state J shows, in addition to the above, a resonance interaction with another identical molecule with $J'=J\pm1$, in whatever vibrational states they may be. It must be noted that both the initial and final levels of an optical transition undergoes resonance if the above condition is satisfied. The calculation of this interaction requires the matrix elements,[33]

[33] R. De L. Kronig and I. Rabi, Phys. Rev. **29**, 262 (1927).
[34] H. Margenau and D. J. Warren, Phys. Rev. **51**, 748 (1937).

$$(J, K, M|\cos\theta|J+1, K, M) = \{[(J+1)^2-K^2][(J+1)^2-M^2]\}^{\frac{1}{2}}/\{(J+1)[(2J+1)(2J+3)]^{\frac{1}{2}}\},$$

$$(J, K, M|\cos\theta|J-1, K, M) = -[(J^2-K^2)(J^2-M^2)]^{\frac{1}{2}}/\{J[2J-1)(2J+1)]^{\frac{1}{2}}\},$$

$$(J, K, M|\sin\theta e^{\pm i\phi}|J+1, K, M\pm1) = \{[(J+1)^2-K^2](J\pm M+1)(J\pm M+2)\}^{\frac{1}{2}}/\{(J+1)[(2J+1)(2J+3)]^{\frac{1}{2}}\}, \quad (9)$$

$$(J, K, M|\sin\theta e^{\pm i\phi}|J-1, K, M\pm1) = \{(J^2-K^2)(J\mp M+1)(J\mp M)\}^{\frac{1}{2}}/\{J[(2J-1)(2J+1)]^{\frac{1}{2}}\}.$$

From Eq. (1) the resonance energy between the molecule with J and $J-1$, respectively, is then given by

$$V_{\text{res}}(J\rightarrow J-1) = \pm\frac{\{(J^2-K^2)(J^2-K'^2)\}^{\frac{1}{2}}}{J^2(2J-1)(2J+1)}[\{(J-M+1)(J-M) \times(J-M')(J+1-M')\}^{\frac{1}{2}}\delta(M, M+1) \times\delta(M', M'-1)+\{(J+M+1)(J+M) \times(J+M')(J+1+M')\}^{\frac{1}{2}}\delta(M, M-1) \times\delta(M', M'+1)-2[(J^2-M^2)(J^2-M'^2)^{\frac{1}{2}} \times\delta(M,M)\delta(M',M')]. \quad (10)$$

The average value vanishes. Its root-mean-square value,

$$\langle V_{\text{res}}^2\rangle^{\frac{1}{2}}(J, K; J-1, K') = \left(\frac{2}{3}\right)^{\frac{1}{2}}\frac{\{(J^2-K^2)(J^2-K'^2)\}^{\frac{1}{2}}}{J\{(2J-1)(2J+1)\}^{\frac{1}{2}}}\frac{\mu^2}{R^3}, \quad (11)$$

may be used as the average value. The resonance interaction energy between the two molecules with J and $J+1$ is obtained by replacing J by $J+1$ in these expressions.

Vibrational resonances may be neglected in microwave spectra because few vibrationally excited molecules are present. If second-order effects are negligible, the interaction energy between symmetric tops may be given, then, by the sum of the above three equations. However, the distribution of the value of J depends on the statistical energy distribution, approximately Maxwell-Boltzmann's, and molecules with J value of maximum population will encounter most frequently molecules of $J\pm1$. In some molecular lines specified by J of small population, the resonance is less effective than the nonresonance, Eq. (7). This is usually realized in the infrared line broadening such as that of HCN.

In line broadening, only interactions relating to the states which can optically combine with one another are influential. Thus, in the inversion spectrum of ammonia, which is emitted by the transition between the states + and −, denoting the symmetry of the wave function with respect to inversion, the antisymmetrical interaction energy is most effective.[35]

b. *Interaction between Linear Molecules*

(*i*) *Linear Molecules with Dipole Moment* μ. (*N_2O, HCN, ICl, ClCN, BrCN, ICN, OCS, etc.*).—The diagonal elements of the dipole interaction vanish because the dipole is pictured as rotating about an axis perpendicular to its electric moment. Other elements are

$$(JMJ'M'|V|jmj'm') = (\mu\mu'/2R^3)\{A(-M)A(M')\delta(m, M-1) \times\delta(m', M'+1)+A(M)A(-M')\delta(m, M+1) \times\delta(m', M'-1)-2B(M)B(M') \times\delta(m, M)\delta(m', M')\}, \quad (12)$$

where

$$A(\pm M)=\left[\frac{(J\pm M+2)(J\pm M+1)}{(2J+3)(2J+1)}\right]^{\frac{1}{2}}\delta(j, J+1) - \left[\frac{(J\mp M+1)(J\mp M)}{(2J+1)(2J-1)}\right]^{\frac{1}{2}}\delta(j, J-1),$$

$$B(M)=\left[\frac{(J+M+1)(J-M+1)}{(2J+3)(2J+1)}\right]^{\frac{1}{2}}\delta(j, J+1) + \left[\frac{(J+M)(J-M)}{(2J+1)(2J-1)}\right]^{\frac{1}{2}}\delta(j, J-1).$$

These are easily obtained from Eqs. (1) and (9) by putting $K=0$.

Clearly, if the two molecules are of the same kind, the interaction becomes first order or second order according to whether there is rotational resonance or not. When a linear molecule with value J encounters another with $J\pm1$, the root-mean-square direction-averaged energy is[20]

$$\langle V_{\text{res}}^2\rangle^{\frac{1}{2}}(J\rightarrow J-1) = \left(\frac{2}{3}\right)^{\frac{1}{2}}\frac{J}{\{(2J-1)(2J+1)\}^{\frac{1}{2}}}\frac{\mu^2}{R^3}. \quad (13)$$

For the case $(J\rightarrow J+1)$, J is replaced by $J+1$ in this expression. For nonresonating collision, London[20,36] calculated the second-order perturbation energy. Denoting by I the moment of inertia of the molecule concerned, one finds

$$V_{\text{dis, Av}} = \frac{2\mu^4 I}{3\hbar^2 R^6}\frac{J(J+1)+J'(J'+1)}{(J+J')(J+J'+2)(J-J'-1)(J-J'+1)}. \quad (14)$$

[35] H. Margenau, Phys. Rev. 76, 1423 (1949).
[36] F. London, Z. Physik 63, 245 (1930).

The value of this energy is positive (repulsive) or negative (attractive) according to the J' of the colliding molecule. Clearly, for $J-J'=\pm 1$ (resonance), Eq. (14) has no meaning.

The ordinary linear molecule has a moment of inertia, I, around 10^{-39} g cm². Thus, from $\hbar^2 J(J+1)/2I = kT$, $J\sim 20$ is most common at room temperature. The thermal energy is usually larger than the difference of successive rotational levels, $\hbar^2 J/I$, especially for lower rotational levels. In this case, where second-order perturbation theory cannot be used, London[26] applied the variational method. Since lower rotational levels are degenerate in the order of thermal energy, the first-order dipole interaction must appear even in the nonresonating case, such as the first-order Stark effect in the hydrogen atom. Hence the eigenvalue is proportional to R^{-3} instead of R^{-6} as in Eq. (14). Keeping this in mind and using the perturbation method, Mizushima[37] took for simplicity the root mean square of the matrix element (12) over M and M':

$$\begin{aligned}(JJ'|V|jj') &= \left(\frac{2}{3}\right)^{\frac{1}{2}} \frac{\mu\mu'}{R^3}\{(2J+1)(2J'+1)\}^{-\frac{1}{2}} \\ &\times\{(J+1)^{\frac{1}{2}}\delta(j,J+1)+J^{\frac{1}{2}}\delta(j,J-1)\} \\ &\times\{(J'+1)^{\frac{1}{2}}\delta(j',J'+1)+J'^{\frac{1}{2}}\delta(j',J'-1)\}. \qquad (15)\end{aligned}$$

Since the rotational quantum number of most colliding molecules is very large, Eq. (15) may be approximated by

$$\begin{aligned}(JJ'|V|jj') = 3^{-\frac{1}{2}}(\mu\mu'/R^3)(2J+1)^{-\frac{1}{2}} &\times\{(J+1)^{\frac{1}{2}}\delta(j,J+1)+J^{\frac{1}{2}}\delta(j,J-1)\} \\ &\times\{\delta(j',J'+1)+\delta(j',J'-1)\}. \qquad (16)\end{aligned}$$

By neglecting energy differences among unperturbed states to get the first-order energy, and considering the fact that the matrix (16) is finite since it is mostly concerned with smaller J's, the energy difference in the transition $J-1\to J$ was taken as

$$0.2\mu\mu'/JR^3, \qquad (17)$$

averaged over J'. But the neglect of the diagonal terms might be very serious, especially for molecules with very small I, in which case the spacing of rotational levels would be rather large. This thermal effect, however, suggests that the R^{-6} law, Eq. (14), goes over to the R^{-3} law, Eq. (17), for small R and, consequently, produces large interactions in the lower rotational levels.

(*ii*) *Nonpolar Linear Molecules.* (O_2, N_2, CO_2, CS_2, *etc.*).—In the interaction between nonpolar molecues, the quadrupole moment interaction may be important in addition to the rotational resonance, if possible, and London's dispersion. This term corresponds to the third one in the expansion (1). The diagonal matrix element is[38]

$$V_{QQ}=\frac{6}{R^5}QQ'[\{K(K+1)-3M^2\}/(2K+3)(2K-1)] \cdot[\{K'(K'+1)-3M'^2\}/(2K'+3)(2K'-1)], \qquad (18)$$

where Q is the quadrupole moment.

If a molecule such as oxygen is considered, Eq. (18) must be revised. Since the electronic ground state of the oxygen molecule is $^3\Sigma$, it has three states, $J=K+1$, K, $K-1$, for each K except for $K=0$ and 1. Using the wave function[39] for these states of oxygen, one obtains the diagonal elements. The factor which comes from one molecule is

$$K\{(K+1)(K+2)-3M^2\}/\{(2K+1)(K+1)(2K+3)\} \quad \text{for} \quad J=K+1, \qquad (19)$$

$$\{K(K+1)-3\}\{K(K+1)-3M^2\}/\{K(K+1)(2K+3)(2K-1)\} \quad \text{for} \quad J=K, \qquad (20)$$

$$(K+1)\{K(K-1)-3M^2\}/\{K(2K+1)(2K-1)\} \quad \text{for} \quad J=K-1. \qquad (21)$$

Again, this interaction is proportional to R^{-5}. Since the mean K is very large at room temperature, the above three terms may be replaced by a single expression,

$$(K^2-3M^2)/4K^2. \qquad (22)$$

Equation (22) may be used for the colliding molecule (primed). Thus, for instance, the energy difference of the radiating molecule with K, M in the transition $K+1\to K$, perturbed by a molecule with K', M', is given by

$$\frac{6}{R^5}QQ'\{(K'^2-M'^2)/4K'^2\}\cdot\{\text{Eq. (19)–Eq. (20)}\}. \qquad (23)$$

The last factor in this formula gives the K dependence. For the transition $K-1\to K$, this factor is replaced by {Eq. (21)–Eq. (20)}.

c. Interaction of a Molecule with a Polarizable Atom

This is a calculation of the induction force. Margenau[40] calculated the interaction energy between a polarizable molecule without μ and a molecule with μ to be given classically without J dependence. Suppose that an axial symmetric molecule has a dipole moment in the direction of z and a quadrupole moment Q, and a polarizable isotropic atom is at a distance R from the molecule, and in the direction θ from the z axis. The electrostatic potential at the atom is

$$U=-\frac{\mu}{R^2}\cos\theta+\frac{1}{2}\left(\frac{3\cos^2\theta-1}{R^3}\right)Q, \qquad (24)$$

[37] M. Mizushima, Phys. Rev. 83, 94 (1951); 84, 363 (1951).

[38] F. London, Z. physik. Chem. (B) 11, 222 (1930).

[39] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, New York, 1935), p. 76.

[40] H. Margenau, Phys. Rev. 49, 596 (1936).

where the quadrupole moment is

$$Q=Q_{zz}-Q_{xx}=\int_{\text{mole}} \rho(z^2-x^2)dv. \quad (25)$$

The origin of coordinates must be chosen at the center of rotation of the molecule. Then, since the field at the polarizable atom is $-\Delta U$, the interaction energy is given by

$$V=-\tfrac{1}{2}\alpha(\Delta U)^2, \quad (26)$$

where α is the polarizability of the perturbing atom. The energy change will be obtained by applying perturbation theory with the wave function of the molecule. As stated before under Sec. *a*, for the inversion spectra of ammonia the antisymmetrical interaction term in Eq. (26),

$$-6\alpha\mu Q/R^7 \cos^3\theta \quad (27)$$

is the most effective for the pressure broadening of the line.[41]

d. Dipole-Quadrupole Interaction

The electrostatic potential energy for an interaction between a molecule with a dipole moment μ_1 and a colliding molecule with a quadrupole moment Q_2 is given by

$$V_{\mu Q}=-\frac{3}{4}\frac{\mu_1 Q_2}{R^4}[\cos\theta_1(1-3\cos^2\theta_2) +2\cos\theta_2 \sin\theta_2 \sin\theta_1 \cos(\phi_1-\phi_2)], \quad (28)$$

where the angles are defined in the same way as in Eq. (1′). With the above expression, Smith and Howard[42] have explained the pressure broadening of the ammonia inversion spectra by foreign nonpolar linear molecules, and determined reasonable values of their quadrupole moments. Smith, Lackner, and Volkor[43] introduced an interaction of this type in addition to the dipole-dipole interaction of Sec. *b*.1 to explain the J and T dependence of the self-pressure broadening of the linear molecules OCS and BrCN.

B. IMPACT THEORY

1. Classical Collision Damping Theory

The shape of spectral lines may be expressed by an intensity distribution function $I(\omega)$ as a function of angular frequency ω. However, in the line broadening problem it is not interesting to find the absolute intensity; rather, attention is paid to the half-width, $\Delta\omega_{\frac{1}{2}}$, intensity maximum frequency, ω_m, and asymmetry of the function associated with a natural frequency ω_0.

$I(\omega)d\omega$ is proportional to the intensity at frequency between ω and $\omega+d\omega$ withdrawn from a light wave of uniform color distribution as it passes an infinitely thin sheet of absorbing molecules. The constant of proportionality is given by the normalization condition,

$$\int_0^\infty I(\omega)d\omega=1. \quad (29)$$

$I(\omega)$ is also proportional to the intensity of emission at ω by a thin layer of the molecules under similar conditions.[2,44] The half-width $\Delta\omega_{\frac{1}{2}}$ is the entire frequency range in which $I(\omega)$ is larger than or equal to half its maximum. Thus, if $I(\omega)$ is symmetrical about its maximum frequency ω_m,

$$I(\omega_m\pm\tfrac{1}{2}\Delta\omega_{\frac{1}{2}})=\tfrac{1}{2}I(\omega_m). \quad (30)$$

Experimentally, it is customary to express ω_m in terms of the shift, $\Delta\omega=|\omega_m-\omega_0|$, of the maximum frequency of the broadened line from its natural frequency ω_0. The effect is referred to as a violet shift or red shift,[2] depending upon the sign of $\omega_m-\omega_0$.

In the case of absorption due to a thin sheet of molecules, the quantity $I(\omega)$ is proportional to the absorption coefficient α at ω and, in practice, the measurement of the line contour $I(\omega)$ itself is the determination of $\alpha=f(\omega)$.

In classical theory, the absorption process is rather obvious. The active atom is replaced by a harmonic oscillator of natural frequency ω_0. Denoting the angular frequency of the incident light wave by ω, we find the equation of motion of the oscillating charge e with mass m under the influence of an electric field $E\cos\omega t$ to be

$$m\left(\frac{d^2x}{dt^2}+\omega_0^2x\right)=eE\cos\omega t, \quad (31)$$

with solution[45]

$$x=\Re\left\{\frac{eEe^{i\omega t}}{m(\omega_0^2-\omega^2)}+C_1e^{i\omega_0 t}+C_2e^{-i\omega_0 t}\right\}. \quad (32)$$

Suppose that a collision takes place at time $t-\theta$. Then, the amplitudes C_1 and C_2 may be affected by the collision. If the collision takes place over an interval of time which is short compared to the period of the oscillation in the impressed field (adiabatic assumption),[45] the transient amplitudes C_1 and C_2 come into play only through their discontinuities at collision and may be determined by the values of x and dx/dt just after the collision. Furthermore, the collision may be assumed to be so strong that the molecule has no memory regarding its orientation or other distribution properties

[41] P. W. Anderson, Phys. Rev. **80**, 511 (1950).
[42] W. V. Smith and R. Howard, Phys. Rev. **79**, 132 (1950).
[43] Smith, Lackner, and Volkor, J. Chem. Phys. **23**, 389 (1955).
[44] J. H. Van Vleck and H. Margenau, Phys. Rev. **76**, 1211 (1949).
[45] J. H. Van Vleck and V. F. Weisskopf, Revs. Modern Phys. **17**, 227 (1945).

before collision (strong collision). This assumption led Lorentz[7] to assume, in addition, that the mean distribution of the oscillators just after collision was not polarized because of the random nature of collisions.

However, since the oscillators are imbedded in an impressed radiation field, there must be an easy direction for them to orient to after collisions and their mean distribution, given by Boltzmann statistics, should be polarized. The Hamiltonian function for this system is

$$H(t)=\frac{p^2}{2m}+\tfrac{1}{2}m(\omega_0 x)^2-exE\cos\omega t,\quad p=m\frac{dx}{dt}.$$

Then, according to Boltzmann statistics, the average value of x after collision must be given by

$$\bar{x}=\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\exp[-H(t-\theta)/kT]dxdp}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp[-H(t-\theta)/kT]dxdp}=\frac{2Ee\cos\omega(t-\theta)}{m\omega_0{}^2},\quad (33a)$$

and

$$\bar{p}=0,\quad \text{at}\quad t_0=t-\theta.\quad (33b)$$

$\bar{x}$ and $\bar{p}$ determine the values of C_1 and C_2, and, finally, give

$$x=\Re\Bigg[\frac{eE\exp[i\omega t]}{m(\omega_0{}^2-\omega^2)}\Bigg\{1-\frac{1}{2}\left(1+\frac{\omega}{\omega_0}\right)\exp[i(\omega_0-\omega)\theta]-\frac{1}{2}\left(1-\frac{\omega}{\omega_0}\right)\exp[-i(\omega_0+\omega)\theta]\Bigg\}+\frac{eE\exp[i\omega(t-\theta)]}{2m\omega_0{}^2}\times(\exp[i\omega_0\theta]+\exp[-i\omega_0\theta])\Bigg].$$

This x must be averaged over θ. In the case of random collisions, the probability that the last collision having taken place before t lies in the time interval between $t-\theta-d\theta$ and $t-\theta$ is $e^{-\theta/\tau}d\theta/\tau$, τ being the mean free time. Averaging over this distribution of θ from $\theta=0$ to ∞, we find x takes the form

$$x=\Re\Bigg[\frac{eE\exp[i\omega t]}{m(\omega_0{}^2-\omega^2)}\Bigg\{1-\frac{(\omega_0+\omega)(\omega/\omega_0{}^2)\tau}{2[(1/\tau)-i(\omega_0-\omega)]}+\frac{(\omega_0-\omega)(\omega/\omega_0{}^2)\tau}{2[(1/\tau)+i(\omega_0+\omega)]}\Bigg\}\Bigg].\quad (34)$$

Writing the square-bracketed term of (35) in the form $Ee^{i\omega t}(x'-ix'')$, we obtain the absorption coefficient[46] α as a function of ω:

$$\alpha=4\pi N^a\omega ex''/c=\frac{2\pi N^a e^2}{mc}\left(\frac{\omega}{\omega_0}\right)^2\times\left(\frac{1/\tau}{(\omega_0-\omega)^2+(1/\tau^2)}+\frac{1/\tau}{(\omega_0+\omega)^2+(1/\tau^2)}\right).\quad (35)$$

This is the absorption line shape. N^a is the number density of absorbing oscillators with natural frequency ω_0. For $\omega_0-\omega\ll\omega_0$, the second term is negligibly small under ordinary conditions, as in the case of atomic spectra or of sharp lines, while in the microwave region, where the width and the frequency are sometimes comparable in order of magnitude, it cannot be overlooked. The Lorentz assumption gives the same formula except for the sign of the second term. Therefore, it predicts a zero effect at low frequency, while it is nearly the same numerically as (35) for high frequencies. If the frequency of oscillation is extremely high, the phase may oscillate rapidly during a collision and thermal equilibrium may not be created after collision. Then, the Lorentz treatment may give a correct answer.

Recently, Bruin[47] reconsidered the collision broadening theory in analogy with the network theory to represent spectral lines graphically in the complex plane.

Equation (35) can be transformed easily into a quantum-mechanical expression if one makes use of the fact that each transition $i\to j$ could be identified with a virtual classical oscillator and applies the rule (see Born,[48] p. 468) of multiplying a classical quantity e^2/m by the oscillator strength

$$f_{ij}=\frac{2m}{3\hbar e^2}\omega_{ij}|\mu_{ij}|^2.$$

Here μ_{ij} is the dipole moment matrix element related with the radiation process $i\to j$; ω_{ij} is the frequency of the radiation. If the incident radiation is weak enough to produce thermal equilibrium, the number densities of the two levels, N_i and N_j, are determined by the Boltzmann distribution law. Since Eq. (35) holds for any possible transition, true absorption or induced emission, and $\omega_{ij}=-\omega_{ji}$, one has

$$\alpha_{ij}(\omega)=\frac{4\pi(N_i{}^e-N_j{}^e)}{3\hbar c}\omega_{ij}|\mu_{ij}|^2\left(\frac{\omega}{\omega_{ij}}\right)^2\times\left(\frac{1/\tau}{(\omega_{ij}-\omega)^2+(1/\tau^2)}+\frac{1/\tau}{(\omega_{ij}+\omega)^2+(1/\tau^2)}\right),\quad (36a)$$

where the superscripts imply thermal equilibrium, and

[46] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 328.

[47] F. Bruin, thesis, Amsterdam (1956).

[48] Max Born, *Optik* (Verlag Julius Springer, Berlin, 1933), p. 441.

$$N_i^e-N_j^e=N_i^e[1-\exp(-\hbar\omega_{ij}/kT)]$$

$$\simeq\begin{cases}\dfrac{\hbar\omega_{ij}}{kT}N_i^e, & \text{for microwaves.}\\ N_i^e, & \text{for the visible region.}\end{cases}$$

For plane polarized radiation, Eq. (36a) should be multiplied by 3. The transition $j\rightarrow i$ involved in Eq. (36a) corresponds to stimulated emission.

Later, Van Vleck and Margenau[44] showed explicitly that the shape of spectral lines for a classical harmonic oscillator and also a Debye slow rotator are the same in absorption and spontaneous emission,[49] provided the energy density obeys the Rayleigh-Jeans law. They showed also that the Lorentz formula is obtained by adding the work done on the molecule by the electromagnetic wave between collisions and at collision. Karplus and Schwinger[50] have given a quantum-mechanical derivation of the Van Vleck-Weisskopf absorption formula (35) and extended it to higher power levels of the exciting radiation, where transitions among the molecular states are included at a rate that is not negligible compared with the collision rate. This treatment invalidates the assumption of thermal equilibrium, although the conditions assumed hardly exist in the visible region.[1] They found that the broadening of the absorption line at high power levels is not to be attributed to any intrinsic modification of the line shape, but rather to a frequency dependent alteration of the populations of the two levels. The net rate of absorption is the difference between the rates of true absorption and stimulated emission, with the common transition probability for the two processes possessing the characteristic frequency dependence $1/[(\omega_{ij}-\omega)^2+(1/\tau^2)]$. The frequency dependence of the population difference may be written

$$\bar{\rho}_{ii}-\bar{\rho}_{jj}=\frac{1/\tau}{(1/\tau)+w}(\rho_i^e-\rho_j^e),$$

where

$$w=\frac{1}{\hbar^2}|V_{ij}|^2\frac{1/\tau}{(\omega_{ij}-\omega)^2+(1/\tau^2)}.$$

$\bar{\rho}_{ii}$ is the mean density of the level i over the collision time θ, and ρ_i^e is that of the isolated molecule at temperature T. V_{ij} is the matrix element of the energy in the external radiation field. Denoting the electric field by $E\cos\omega t$ and the induced dipole moment by μ, one finds $V_{ij}=\mu_{ij}E$. Then, the power absorbed per unit volume is

$$P_a=\tfrac{1}{2}\omega\hbar\omega(\bar{\rho}_{ii}-\bar{\rho}_{jj})N^a,$$

where the transition probability for absorption or stimulated emission has been placed equal to $(1/2)w$, since both processes contribute equally to the rate at which radiative transitions tend to decrease the population difference. If P denotes the radiation power incident on the gas per unit time per unit area, $P=E^2c/8\pi$, c being the velocity of light. Hence,

$$P_a=\frac{4\pi\omega P}{c\hbar}|\mu_{ij}|^2\frac{(1/\tau)(N_i^e-N_j^e)}{(\omega_{ij}-\omega)^2+(1/\tau^2)+\dfrac{8\pi P}{c\hbar^2}|\mu_{ij}|^2}. \quad (36b)$$

Thus, as P increases, the value of P_a saturates. In the limit of complete saturation one has

$$P_a=\frac{\hbar\omega}{2\tau}(N_i^e-N_j^e)\simeq\frac{\hbar^2\omega}{2\tau}\frac{\omega_{ij}N_j^e}{kT}.$$

This saturation effect, Eq. (36b), has been experimentally confirmed with ammonia in the microwave region.[51]

The half-width of the line shape may be given by

$$\Delta\omega_{\frac{1}{2}}=\frac{2}{\tau}=2\pi\rho^2\bar{v}N \quad (37)$$

for the case of resonance absorption in Eq. (35). In this ρ is the optical collision diameter, related to the force law at collision, $\bar{v}$, the rms velocity of impact, and N the number of colliding atoms per unit volume. The line shift, in this approximation, is obviously zero. The linear dependence of the half-width on N is characteristic of this treatment. The collision mechanism replaced by the assumption of strong collision has lost the chance to include the force law precisely in the theory. However, the magnitude of the optical collision diameter corresponds to that of the intermolecular force concerned. The optical collision diameter[48] is usually greater by a factor two or three, sometimes by a factor one hundred, than the gas-kinetic one which is related to momentum transfer. (See Table III.) This discrepancy may be partly attributed to the inclusion of excited states in optical collision diameters, but it makes the theoretical treatment easier, since in such cases the flying path of perturbers is approximated by a straight line.

2. Classical Fourier Integral Theory

Assume a radiating classical oscillator with natural frequency ω_0. Although the frequency of the oscillator would be ω_0 again after completion of an adiabatic collision, the change of the oscillation during collision would be so appreciable that the phase would be quite

[49] About the line shape modified by self-absorption, if any, in emission, see R. D. Cowan and G. H. Dieke, Revs. Modern Phys. 20, 418 (1948). Especially, in arc this effect should be taken into consideration; see H. Edels and J. D. Craggs, Proc. Phys. Soc. (London) A64, 562 (1951), or in flame; see Huldt and Knoll (reference 12).

[50] R. Karplus and J. Schwinger, Phys. Rev. 73, 1020 (1948). Also, see P. I. Richards and H. S. Snyder, Phys. Rev. 73, 269 (1948).

[51] T. A. Pond and W. F. Cannon, Phys. Rev. 72, 121 (1947); B. Bleaney and R. P. Penrose, Proc. Phys. Soc. (London) 60, 83 (1948); R. L. Carter and W. V. Smith, Phys. Rev. 73, 1053 (1948); Townes, reference 179.

different from what the oscillator would have if the collision had not taken place. The phase shift of the oscillator is given by

$$\Theta = \int_{\Delta\tau} \Delta\omega(\tau) d\tau, \tag{38}$$

where $\Delta\tau$ is the collision duration and $\Delta\omega(\tau)$ the difference between the frequencies between and at collisions. Since an adiabatic collision and no change of amplitude are assumed, it may affect the radiation only through Θ, regardless of the manner by which it is produced. The strong collision assumed in the previous section is equivalent to the assumption that Θ is arbitrarily large. But a number of small phases could add up to a large phase shift and affect a spectral line. The main purpose of the modified theory described here, however, is to account for the magnitude of $\Delta\omega(\tau)$ resulting from disturbances at collision.

If the frequency of the oscillator changes with time under the influence of colliding molecules, the oscillation may be expressed by

$$x(t) = x_0 \exp\left[-i\int_0^t \omega(t')dt'\right]. \tag{39}$$

The Fourier analysis of this disturbed oscillation may give information on $I(\omega)$. According to classical mechanics, the power radiated by a charge e oscillating in one direction is $(2/3)e^2c^{-3}(d^2x/dt^2)^2$. Hence, taking Fourier components of the displacement (39) and multiplying it by ω^2 to obtain acceleration, the power emitted in the spectral ω and $\omega+d\omega$ is

$$I(\omega)d\omega = \frac{e^2\omega^4}{3\pi c^3}x_0^2\left|\int_{-\infty}^{\infty} dt \times\exp\left\{-i\int_0^t \omega(t')dt' + i\omega t\right\}\right|^2 d\omega. \tag{40}$$

This formula is what Weisskopf[1] obtained by the WKB method and later by many people in a more general way.[52,53]

To reduce the general formula (40) to a more comprehensive form approximations may be made. For instance, if the Lorentz view is followed, the oscillation stops completely at strong collision, and the wave train emitted is finite in its length, say, from $t=0$ to θ. Then, the limits of integration are from 0 to θ. After averaging $I(\omega)$ over the distribution of θ as before, the formula (35) is obtained, except for unimportant factors (see Born,[48] p. 437). This is true also if the recalculation is carried out by assuming the Rayleigh-Jeans law and the equipartition value for the statistical average of the square of the amplitude of a harmonic oscillator. Since there is no impressed field, the consideration of Boltzmann statistics is unnecessary.

The phase shift Θ may be supposed to operate only as a temporary interrupter of the radiation process. Although the distribution of the magnitude of Θ is continuous, Lenz,[54] Weisskopf,[1] and others[55] assumed that the effective phase shift must be greater than some critical value—unity in Weisskopf's treatment, $\pi/2$ or 1.1 in others. The Lorentz view requires that the line intensity be lowered inversely proportionally to the number of collisions per unit time; the improved view does not mean the complete stoppage of the emission of radiation, which agrees with many experiments.

Although, by critical values of effective phase shift, early investigators determined theoretically the collision diameter, the cut-off distance of the radiation process, there was no definite meaning attached to the magnitude of the effective phase shift. Reinsberg, Lindholm[56] and Foley[22] improved this point, including every magnitude of phase shift. Rewriting the formula (40), Foley obtained

$$I(\omega) = \frac{e^2\omega^4}{3\pi c^3}x_0^2\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \times\exp\left[-i\int_{t_2}^{t_1} \omega(t')dt' + i\omega(t_1-t_2)\right]. \tag{41}$$

The substitution $t_1-t_2=\tau$, $t_2=t_0$ leads to

$$I(\omega) = \frac{2e^2\omega^4}{3\pi c^3}x_0^2\Re\int_0^{\infty} d\tau \int_{-\infty}^{\infty} dt_0 \times\exp\left[-i\int_{t_0}^{t_0+\tau} \omega(t')dt' + i\omega\tau\right]. \tag{42}$$

Denoting the natural frequency of the oscillator by ω_0, and putting

$$\Delta\omega(t') = \omega(t') - \omega_0, \quad \Delta\omega = \omega - \omega_0, \tag{43}$$

where $\Delta\omega$ is measured from the unshifted line maximum, we find the intensity distribution is given by

$$I(\omega) = \frac{2e^2\omega^4}{3\pi c^3}x_0^2\Re\int_0^{\infty} d\tau \int_{-\infty}^{\infty} dt_0 \times\exp\left[-i\int_{t_0}^{t_0+\tau} \Delta\omega(t')dt' + i\Delta\omega\tau\right]. \tag{44}$$

This formula may be thought of as the Fourier ampli-

[52] P. W. Anderson, Phys. Rev. **76**, 647 (1949). Furthermore, the fact that most of the optical collisions take place without momentum transfer may allow this assumption.

[53] H. Margenau and R. Meyerott, Astrophys. J. **121**, 194 (1955).

[54] W. Lenz, Z. Physik **80**, 423 (1933).

[55] H. Kallman and F. London, Z. physik. Chem. **B2**, 207 (1929); H. Kuhn and F. London, Phil. Mag. (7) **18**, 983 (1934); G. Burkhardt, Z. Physik **115**, 592 (1940).

[56] C. Reinsberg, Z. Physik **111**, 95 (1938); E. Lindholm, Arkiv Mat. Astron. Fys. **28B**, 1 (1942).

tude of the function

$$\varphi(\tau)=\int_{-\infty}^{\infty}dt_0\exp\left[-i\int_{t_0}^{t_0+\tau}\Delta\omega(t')dt'\right],\qquad(45)$$

which is the average value of the phase difference occurring between the time $t_0+\tau$ and t_0 over the complete interval of t_0, that is, the radiation process. Recently the general formula (44) has been used to obtain a line shape at ordinary pressures.

If we assume an adiabatic collision, only the phase shift at collision is important. Since the collision duration is infinitely short, each collision is binary and independent. Therefore, the phase difference appearing in the above integral is approximated by the sum of such phase shifts taking place in the time interval.

$$\exp\left(-i\int_{t_0}^{t_0+\tau}\Delta\omega(t')dt'\right)=\exp(-i\sum_{i=0}^{n}\Theta_i).\qquad(46)$$

This must be averaged over the distribution of n. If the mean free time between random collisions is τ_0, the probability that n collisions take place in the interval τ is $(\tau/\tau_0)^n(e^{-\tau/\tau_0}/n!)$. In this form of interpretation of the above integral (45), the average over t_0 may be converted to a spatial average. If we consider one oscillator for the complete interval t_0 we obtain the same result as that found by considering a great number of oscillators in a short time interval. If the latter view is followed, the average over t_0 may turn out to be the average of Θ_i over its distribution. Two types of averaging processes are involved, and the order in which they are carried out does not affect the result.[37] Denoting the probability of occurrence of a phase shift Θ within an interval between Θ and $\Theta+d\Theta$ by $\rho(\Theta)d\Theta$, we can write the above integral (45) as

$$\sum_{n=0}^{\infty}(\tau/\tau_0)^n\left(\frac{e^{-\tau/\tau_0}}{n!}\right)\prod_{i=1}^{n}\int_{-\infty}^{\infty}d\Theta_i\rho(\Theta_i)e^{-i\Theta_i},\qquad(47)$$

since the phase shift distribution from $-\infty$ to ∞. The $\rho(\Theta)$ may be understood as the collision parameter producing a phase shift Θ, if we assume straight paths for colliding molecules. Since the value of the integral with respect to Θ_i is the same for any suffix i, it may be written $A-iB$.

$$A=\int_{-\infty}^{\infty}d\Theta\rho(\Theta)\cos\Theta,\quad B=\int_{-\infty}^{\infty}d\Theta\rho(\Theta)\sin\Theta.\qquad(48)$$

Hence, (45) is reduced to

$$\varphi(\tau)=\exp\left[\frac{A-iB}{\tau_0}\tau-\frac{\tau}{\tau_0}\right].\qquad(49)$$

Thus, the intensity formula turns out to be

$$I(\Delta\omega)=\frac{2\omega^4}{3\pi c^3}(ex_0)^2\frac{(1-A)/\tau_0}{\{(1-A)/\tau_0\}^2+\{B/\tau_0-\Delta\omega\}^2}.\qquad(50)$$

The correlation function method,[6,52] when one does not assume a distribution of n, gives the same result as above for Eq. (44).

According to Eq. (50), the intensity maximum occurs at $\Delta\omega=B/\tau_0$, so that if B is positive the line center shifts to the blue, and, if negative, to the red. The half-width is given by $2(1-A)/\tau_0$ with no assymmetry. According to the definitions of A and B, and noting that the distribution function $\rho(\Theta)$ is normalized, we observe that the half-width and line shift are given by

$$\Delta\omega_{\frac{1}{2}}=2\int\rho(\Theta)(1-\cos\Theta)d\Theta/\tau_0.$$

and

$$\Delta\omega_m=\int\rho(\Theta)\sin\Theta d\Theta/\tau_0.\qquad(51)$$

Since $\sin\Theta$ is an odd function, and if positive as well as negative values of Θ are equally probable, the shift of the maximum frequency is zero, as in the case of resonance broadening. The function $\rho(\Theta)d(\Theta)/\tau_0$ can be interpreted as the number of collisions per unit time with phase shift between Θ and $\Theta+d\Theta$. Thus, both the width and shift are proportional to density, the number of molecules per unit volume.

There are many cases where the molecular energy under the influence of another molecule at a distance R is expressed by

$$(E_0+\hbar\gamma/R^p)_i,$$

where i refers to the initial as well as to the final state of the radiative transition concerned. Hence,

$$\Delta\omega(t')=\Delta\gamma/R(t')^p,\qquad(52)$$

$\Delta\gamma$ being the difference of the force constants for the two states. If the path of a colliding molecule with relative velocity v is straight, as is probable in the case of no complete interruption of the radiation process,[52] $R=(\rho^2+v^2t'^2)^{\frac{1}{2}}$ with impact parameter ρ and time t'. With the help of Eq. (52) the phase shift, Eq. (38), is, under the binary interaction, given by[56a]

$$\Theta=\int_{-\infty}^{\infty}\Delta\gamma/(\rho^2+v^2t'^2)^{p/2}dt'$$

$$=\left(\frac{\Delta\gamma}{v\rho^{p-1}}\right)\pi\Gamma(p-1)2^{2-p}\left[\Gamma\left(\frac{p}{2}\right)\right]^{-2}.\qquad(53)$$

The extension of integration limits to infinity may not be inconsistent with the assumption of adiabatic collision.

The number of collisions with relative velocity between v and $v+dv$ and impact parameter between ρ and $\rho+d\rho$ can be obtained easily by the kinetic theory of gases:

$$\rho(\Theta)d\Theta/\tau_0=8\pi^{\frac{1}{2}}(m/2kT)^{\frac{3}{2}}\exp(-mv^2/2kT)v^3dv\rho d\rho gN,$$

[56a] D. B. Dettaan, *Nouvelles Tables d'Intégrales I éfines* (P. Engels, Libraire Editeur, Leide, 1867), p. 69.

where m is the reduced mass, and N the number of molecules per unit volume. The term g refers to the fact that the force law depends on the various states of colliding molecules and is the probability of a molecule being in a particular rotational state; it must be summed over all states.

Thus, in the case of the force law of inverse pth power, the half-width and line shift are

$$\Delta\omega_{\frac{1}{2}} = 2\pi^{(p+3)/(2p-2)}(kT/2m)^{(p-3)/(2p-2)}$$
$$\times\Gamma\left(\frac{2p-3}{p-1}\right)\Gamma\left(\frac{p-3}{p-1}\right)$$
$$\times\left[\Gamma(p-1)\Big/\left\{\Gamma\left(\frac{p}{2}\right)\right\}^2\right]^{2/(p-1)}$$
$$\times\sin\left\{\frac{p-3}{2p-2}\pi\right\}\langle|\Delta\gamma|^{2/(p-1)}\rangle N, \quad (54a)$$

$$\Delta\omega_m = \tfrac{1}{2}\tan\left(\frac{\pi}{p-1}\right)\Delta\omega_{\frac{1}{2}}\langle\Delta\gamma/|\Delta\gamma|\rangle. \quad (54b)$$

In order that these expressions be convergent, p must be larger than 2. The bracket $\langle\ \rangle$ implies an average over g. The value of $\langle|\Delta\gamma|^{2/(p-1)}\rangle$ may be sometimes approximated by the mean square value. For some special values of p, the equation gives

$$\Delta\omega_{\frac{1}{2}} = 2\pi^2\langle|\Delta\gamma|\rangle N, \quad \text{for} \quad p=3, \quad (55)$$

$$\Delta\omega_{\frac{1}{2}} = 13.4(kT/2m)^{1/4}\langle|\Delta\gamma|^{1/2}\rangle N, \quad \text{for} \quad p=5, \quad (56)$$

$$\Delta\omega_{\frac{1}{2}} = 12.9(kT/2m)^{3/10}\langle|\Delta\gamma|^{2/5}\rangle N, \quad \text{for} \quad p=6, \quad (57)$$

respectively. Lindholm[57] derived a formula essentially equivalent to Eq. (57).

It is interesting to consider the range of values of phase shifts Θ which affect the line shape the most. Under the force law, Eq. (52), $\rho(\Theta) = C\Theta^{-1/(p-1)}$ in Eq. (51). Then, when the sign of Θ is positive at all collisions, the integrand of $\Delta\omega_m$ in Eq. (51) oscillates between positive and negative with slowly varying amplitude in the range of $\Theta > 1$. Hence, the integrand in this range, corresponding to strong phase shifts, averages nearly to zero and contributes very little to the value of $\Delta\omega_m$. Therefore, the line shift should come from the range of $\Theta < 1$. This range is the one completely neglected, by Weisskopf[1] and others,[54,55] as ineffective on line shape. On the other hand, the greater part of the value of $\Delta\omega_{\frac{1}{2}}$ in Eq. (51) comes from the range of $\Theta > 1$, since the integrand of $\Delta\omega_{\frac{1}{2}}$ is always positive. Thus the older view of Weisskopf[1] and others[54,55] on line broadening using the idea of critical effective phase shifts can be well understood. However, the result, obtained here, that the strong phase shift averages to zero, still leaves some ambiguity in the theory. An analysis[58] shows that the statistical theory in Section C is better for such collisions. Actually, the Fourier integral theory is good for a line shape near its center.[54]

If the sign of $\Delta\gamma$ is common to all collisions, Eq. (54) gives Foley's relation

$$\Delta\omega_{\frac{1}{2}}/|\Delta\omega_m| = 2\cot\left(\frac{\pi}{p-1}\right). \quad (58)$$

The above condition is valid for heavier atoms perturbed by foreign gases where van der Waals forces are important. Although the observed values of line shifts and widths in atomic lines perturbed by foreign gases give widely distributed values for the ratio, Foley[22] shows that p in (58) is nearly 6, assuming the predominance of only one type of force. (See Table III.) However, in most cases of molecular spectra the above condition does not hold and Eq. (58) cannot be used.

Since the interaction forces which are predominant for molecules at low pressures are rather well known, the test of Eqs. (54a) and (54b) has been made with good over-all agreement with observations. For atomic lines the test can hardly be made directly, except for the linearity relationship to N as well as the test of (58). The tests for both microwave and atomic lines require quite different approaches, and may call for separate experimental justifications.

3. Quantum-Mechanical Fourier Integral Theory

In the standard Fourier integral theory, as shown in the previous section, the inclusion of all phase shifts at collisions removed the ambiguous collision diameter and gave rise to a line shift. However, its basic assumption is that the collision is adiabatic and does not change the amplitude of the oscillator in a particular direction.

If molecules have rotational levels whose spacing is small in comparison with the thermal energy, transitions between levels (reorientation of the molecule) may be induced even by adiabatic collisions (infinitely short collision duration) and the amplitude will change. As has been considered by Spitzer,[59] Van Vleck and Weisskopf,[45] Fröhlich,[60] and later under more refined assumptions for low frequency limiting cases by Gross,[61] this effect places the results of the old Lorentz theory in contradiction to the theory of dielectric relaxation developed by Debye[62] and others,[63] as shown in Sec. B1. The relaxation theory of dielectrics treats broadening as the limiting case of a spectral line of zero frequency,

[57] E. Lindholm, Arkiv Mat. Astron. Fys. 32A, 17 (1945).

[58] T. Holstein, Phys. Rev. 79, 744 (1950). Also, see reference 37.

[59] L. Spitzer, Jr., Phys. Rev. 58, 348 (1940).

[60] H. Fröhlich, Nature 157, 478 (1946); *Theory of Dielectrics* (Oxford University Press, New York, 1949). This model for one degree of freedom was extended to the case of two degrees of freedom by R. Huby, Rept. Brit. Elect. Allied Industr. Research Assoc. [reference L/T 179 (1947)].

[61] F. P. Gross, Phys. Rev. 97, 395 (1955).

[62] P. Debye, *Polar Molecules* (Chemical Catalog Company, New York, 1929), Chap. V.

[63] W. Kauzmann, Revs. Modern Phys. 14, 12 (1942).

in which case the standard Fourier integral theory also does not give any pressure effect. Thus, it is necessary to develop a generalized Fourier integral theory.

The starting point is given by the radiation theory of quantum mechanics. This method was followed by Lindhold,[57] Foley,[22] Anderson,[52,64] Mizushima,[37] Karplus and Schwinger,[50] Leslie,[65] and Bloom and Margenau.[66] It is customary to treat the intermolecular motion classically. Since the range of the intermolecular force, which is important to line broadening in the valid range of the Fourier integral theory, is several times larger than the wave packet associated with a moving molecule, the packet may be considered as a point without a large error, except for very light molecules such as helium. The coordinates of a colliding molecule are known functions of time, and the molecular interaction is expressed in terms of a definite time. Thus, the following discussions are limited to the classical region well outside of an absorber. First a general formula will be obtained, and then it will be applied to the low-pressure case when multiple collisions can be neglected.

The Hamiltonian of a molecule undergoing perturbations $C(t)$ due to collisions and $F(t)$ due to light waves will be

$$H(t)=H_0+C(t)+F(t), \tag{59}$$

where H_0 is the Hamiltonian of an unperturbed molecule in a stationary state denoted by φ_m. This state is usually highly degenerate, but has a discrete energy. Thus one can know which spectral line is under consideration. While Mizushima[37] solved the Schrödinger equation with this H under the assumption of adiabatic hypothesis (no transition between atomic states due to collision) following Foley's method, Bloom and Margenau[66] divided the transitions concerned into two parts, one due to collisions and another due to the radiation field. They considered only the transitions relating to the radiation process. The wave function of the molecule perturbed by collisions will be expressed in terms of φ_m by

$$\Phi_i(t)=\sum_m U_{mi}(t)\varphi_m, \tag{60}$$

which satisfies the Schrödinger equation $i\hbar\partial\Phi_i/\partial t=[H_0+C(t)]\Phi_i$, with an initial condition $\Phi_i(t_0)=\varphi_i$, i.e. $U(t_0)=1$. Since this Φ_i forms a complete orthogonal set at every instant, the solution of the Schrödinger equation with the Hamiltonian H (59) can be expanded as $\Psi=\sum_i a_i(t)\Phi_i(t)$. In this $a_i(t)$ is the probability amplitude of the collision-smeared state Φ_i. The density of the ith level at t is $\rho_i(t)=a_i^*(t)a_i(t)$. This transition probability amplitude is induced by the light wave and is related to the radiation problem.

The amplitudes are subject to the equation[67]

$$i\hbar da_k/dt=\sum_i F_{ki}a_i, \tag{61}$$

with

$$F_{ki}=\int \Phi_k^*F\Phi_i dv.$$

The increasing rate of the population in the kth level due to absorption depends on the value of $|F_{ki}|^2$ and the population in the ith level. However, there must be induced radiations, which are connected to the transition $k\rightarrow i$ leading to the decrease of $\rho_k(t)$. Thus,

$$\rho_k(t)=\rho_k(t_0)+\sum_i \rho_i(t_0)\left|(i\hbar)^{-1}\int_{t_0}^{t} F_{ki}dt\right|^2 - \sum_i \rho_k(t_0)\left|(i\hbar)^{-1}\int_{t_0}^{t} F_{ki}dt\right|^2_0.$$

Bloom and Margenau[66] solved Eq. (61) rigorously and obtained the $\rho_k(t)$ up to second order in F_{ki}, which is the same as above.

The F_{ki} can be expressed in the stationary states base φ.

$$F_{ki}=\sum_{ml} U_{mk}^*U_{li}\int \varphi_m^*F\varphi_l dv.$$

The explicit form of the field perturbation in the dipole approximation is $F(t)=-\mathbf{E}\cdot\boldsymbol{\mu}\cos(\omega t+\alpha)$. When one sets $\mathbf{E}\cdot\boldsymbol{\mu}=E\mu\cos\theta$ and averages over θ and over the random phase angle α, one obtains

$$\rho_k(\infty)-\rho_k=\frac{E^2}{12\hbar^2}\sum_i(\rho_i-\rho_k)\left\{\left|\int_{-\infty}^{\infty}dt\mu_{ki}(t)e^{i\omega t}\right|^2+\left|\int_{-\infty}^{\infty}dt\mu_{ki}(t)e^{-i\omega t}\right|^2\right\}, \tag{62}$$

where

$$\mu_{ki}(t)=\sum_{ml} U_{mk}^*(t)U_{li}(t)\int \varphi_m^*\mu(t_0)\varphi_l dv. \tag{63}$$

Hence, the intensity distribution may be expressed by

$$I(\omega)=\hbar\omega(\rho_k(\infty)-\rho_k)=\frac{2\pi\omega I_0(\omega)}{3\hbar c}\sum_i(\rho_i-\rho_k)\times\left\{\left|\int_{-\infty}^{\infty}dt\mu_{ki}(t)e^{i\omega t}\right|^2+\left|\int_{-\infty}^{\infty}dt\mu_{ki}(t)e^{-i\omega t}\right|^2\right\}, \tag{64}$$

where $I_0(\omega)$ represents the incident light beam. Equation (64) has been also obtained by the calculation of work done by the light wave.[66] This result suggests a possible process of separation of $I(\omega)$ into lines $I_{ki}(\omega)$.

[64] P. W. Anderson, dissertation, Harvard University, 1949 (unpublished).
[65] D. C. M. Leslie, Phil. Mag. 42, 37 (1951).
[66] S. Bloom and H. Margenau, Phys. Rev. 90, 791 (1953).
[67] L. I. Schift, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), pp. 190–193.

When $C(t)$ is absent, $\mu_{ki}(t)=\mu_{ki}(0)\exp(i\omega_{ki}t)$, where $\omega_{ki}=(E_k-E_i)/\hbar$. Therefore, one may take it as an approximate solution in the case of the weak collision perturbation. Hence, the first integrand in Eq. (64) has approximately with exponential factor $\exp i(\omega_{ki}+\omega)t$, the second the factor $\exp i(\omega_{ki}-\omega)t$. If $E_k>E_i$ the second of these produces resonance and the first can be neglected; otherwise their role is reversed. In that case,

$$\rho_i\left|\int_{-\infty}^{\infty}dt\mu_{ki}(t)e^{-i\omega t}\right|^2 \qquad (65)$$

represents absorption, and

$$-\rho_k\left|\int_{-\infty}^{\infty}dt\mu_{ki}(t)e^{-i\omega t}\right|^2 \qquad (66)$$

represents induced emission.

The problem is reduced, then, to obtain the dipole moment $\mu_{ki}(t)$. It is well known that it is possible to rewrite Eq. (63) in the vector form,[68]

$$\mu(t)=U^{-1}(t)\mu(t_0)U(t), \qquad (67)$$

where $U(t)$ is unitary and satisfies the equations

$$i\hbar dU/dt=(H_0+C(t))U, \qquad (68)$$

and

$$U(t_0)=1.$$

If one sets $U(t)=\exp(H_0t/i\hbar)T(t)$, where $T(t)$ is unitary, the differential equations lead to

$$i\hbar dT/dt=C'(t)T(t), \qquad (69)$$

where

$$C'(t)=\exp\left(-\frac{H_0t}{i\hbar}\right)C(t)\exp\left(\frac{H_0t}{i\hbar}\right).$$

The matrix element of $C'(t)$ is given by

$$C_{mm'}'(t)=C(t)_{mm'}\exp[(E_{m'}-E_m)t/i\hbar]. \qquad (70)$$

Then,

$$\mu_{ki}(t)=\sum_{ml}(T(t)_{km}{}^{-1})\exp(-i\omega_{ml}t)\mu_{ml}(t_0)(T(t)_{li}). \qquad (71)$$

The diagonal elements of $T(t)$ give the phase shifts of $\mu(t)$ due to collision interaction $C(t)$, while the off-diagonal elements—or their squared absolute values—gives the transition probabilities due to this interaction.

Because $C(t)$ and H_0 do not commute, $C'(t)$ defined above does not commute with itself at all times. However, in cases where $C(t)$ is either diagonal already, or can be made so in a simple manner such as replacing the so-called van der Waals forces, a second-order phenomenon, by a first-order effective Hamiltonian, Eq. (69) integrates easily.

$$T_{kk}(t)=\exp\left[\frac{1}{i\hbar}\int^t C_{kk}(t')dt'\right] \qquad (72)$$

[68] M. Born and P. Jordan, *Elementare Quantenmechanik* (Verlag Julius Springer, Berlin, 1930), p. 327.

Thus, (71) leads to

$$\mu_{ki}(t)=\mu_{ki}(t_0)\exp\left\{-i\omega_{ki}t-\frac{1}{i\hbar}\times\left(\int^t C_{kk}(t')dt'-\int^t C_{ii}(t')dt'\right)\right\}. \qquad (73)$$

The substitution from (73) into (65) gives the classical Fourier integral formula (40) for absorption, if we assume that ρ_k (66) is negligibly small. In Eq. (73) the term,

$$\frac{1}{\hbar}\left(\int^t C_{kk}(t')dt'-\int^t C_{ii}(t')dt'\right),$$

is equivalent to the phase shift, Eq. (53), at a collision.

For a binary case, if one assumes a straight path for a colliding molecule, Eq. (72) leads to

$$T_{kk}(\rho)=\exp\left[-i\int_{-\infty}^{\infty}C_{kk}([\rho^2+v^2t'^2]^{\frac{1}{2}})/hdt'\right], \qquad (74)$$

where ρ is the impact parameter. Of course, T may be a function of angles, in addition to ρ, required to specify one type of the collision. However, because of the random collision the effect of collisions may be isotropic, so that a collision can be represented by ρ on the average. Hence, Eq. (73) reduces Eq. (65) to the standard Fourier integral theory. From the interaction Hamiltonian $C(t)$, $\Delta\gamma$ defined by Eq. (52) can be obtained. Thus $\Delta\omega_{\frac{1}{2}}$ or $\Delta\omega_m$ can be evaluated by Eq. (54). Since $C(t)$ is generated by a classical thermal motion of a perturber, it cannot be greater than the thermal energy per particle of order kT. Thus any broadening, such that $\hbar\Delta\omega>kT$, is out of the present treatment.

This procedure is generalized to include nondiagonal terms as follows. By a unitary transformation A, $C'(t)$ is diagonalized.

$$AC'(t)A^{-1}=V_{ij}\delta_{ij}. \qquad (75)$$

Transforming by A, Eq. (69) becomes

$$i\hbar A\frac{dT}{dt}A^{-1}=V_{ij}\delta_{ij}AT(t)A^{-1}.$$

If dA/dt is neglected by the adiabatic assumption (very slow thermal motion), we obtain

$$T(t)=A^{-1}\left\{\exp\frac{1}{i\hbar}\int^t V_{ij}dt\right\}\delta_{ij}A. \qquad (76)$$

From Eqs. (65), (71), and (76), the intensity distribution can be obtained.

Sometimes the transformation A in Eq. (75) may be

difficult to obtain. In this case, by introducing higher order terms into $C(t)$ and ignoring the noncommuting property of $C'(t)$ with itself, the solution of Eq. (69) will be approximated by Eq. (72) or, for independent binary collisions, by Eq. (74). If the value of the integral,

$$\int_{-\infty}^{\infty} C'_{mm'}((\rho^2+v^2t'^2)^{\frac{1}{2}})/\hbar dt' = P_{mm'}(\rho), \qquad (77)$$

is small (weak collisions), Eq. (74) can be expanded as

$$T_{mm'}(\rho) = \delta_{mm'} - iP_{mm'}(\rho) + P_{mm'}{}^2(\rho)/2 - \cdots. \qquad (78)$$

This gives a method of successive approximations.[52]

Since collisions are independent, $T(\rho)$ operates independently on the degenerate levels (m, m', m'', m''') belonging to the initial (k) and final (i) eigenvalues. If n collisions have taken place during a time interval τ, from Eq. (67) $\mu_{ik}(t_0)$ transforms to

$$\mu_{ik}'(t_0+\tau) = T_i^{-1}(\rho_n)T_i^{-1}(\rho_{n-1})\cdots \times T_i^{-1}(\rho_1)\mu_{ik}'(t_0)T_k(\rho_1)\cdots T_k(\rho_n),$$

where $\mu' = \exp(-H_0t/i\hbar)\mu \exp(H_0t/i\hbar)$. It is easy to introduce the degeneracy into this expression:

$$\mu_{ik}'(t_0+\tau)_{m'm} = \sum_{m''m'''} \{\prod_j T_i^{-1}(\rho_j)\}_{m'm''}\mu_{ik}'(t_0)_{m''m'''} \times\{\prod_j T_k(\rho_j)\}_{m'''m}. \qquad (79)$$

Because the effect of collisions is isotropic, $\mu_{ik}'(t_0+\tau)$ and $\mu_{ik}'(t_0)$ have exactly the same transformation properties under rotation. Hence, this can be written as

$$\mu_{ik}'(t_0+\tau)_{m'm} = \mu_{ik}'(t_0)_{m'm}F(\tau) \qquad (80)$$

with a scalar function $F(\tau)$.

If we follow the same procedure as followed in the derivation of Eq. (42), Eq. (65) can lead to

$$I_{ki}(\omega) = \rho_i \int d\tau e^{i(\omega-\omega_{ki})\tau}\varphi_{ki}(\tau), \qquad (81)$$

with

$$\varphi_{ki}(\tau) = \langle \sum_{m'} \mu_{ki}'^*(t_0)_{mm'}\mu_{ik}(t_0+\tau)_{m'm}\rangle_{\text{Av over } t_0}. \qquad (81a)$$

Equation (45) and Eq. (81a) are equivalent. If an elementary time interval $d\tau$ is chosen long compared to the duration of one collision which will be represented by $T(\rho)$, one has

$$\delta\varphi_{ki}(\tau) = \varphi_{ki}(\tau+d\tau) - \varphi_{ki}(\tau) = \langle \sum_{m'm''m'''} \mu_{ki}'^*(t_0)_{mm'}[T_i^{-1}(\rho)_{m'm''}\mu_{ik}'(t_0)_{m''m'''}T_k(\rho)_{m'''m} - \mu_{ki}'(t_0)_{m'm}]\rangle F(\tau). \qquad (82)$$

On the other hand, Eq. (80) gives

$$\delta\varphi_{ki}(\tau) = \langle \sum_{m'} \mu_{ki}'^*(t_0)_{mm'}\mu_{ik}'(t_0)_{m'm}\rangle dF(\tau). \qquad (83)$$

Therefore, it follows that

$$\frac{dF(\tau)}{d\tau} = \left\langle \frac{\sum_{m'm''m'''} \mu_{ki}'^*(t_0)_{mm'}T_i^{-1}(\rho)_{m'm''}\mu_{ik}'(t_0)_{m''m'''}T_k(\rho)_{m'''m}}{\sum_{m'} \mu_{ki}'^*(t_0)_{mm'}\mu_{ik}'(t_0)_{m'm}} - 1 \right\rangle F(\tau).$$

The average over t_0 is replaced by an average over ρ. The probability that a collision of the type ρ occurs in $d\tau$ is given by $Nvd\tau 2\pi\rho d\rho$, where N is the number density of perturbers and v the mean velocity. Therefore,

$$dF(\tau)/d\tau = -Nvd\tau F(\tau)\int 2\pi\rho d\rho S(\rho), \qquad (84)$$

where

$$S(\rho) = 1 - \frac{\sum_{m'm''m'''} \mu_{ki}'^*(t_0)_{mm'}T_i^{-1}(\rho)_{m'm''}\mu_{ik}'(t_0)_{m''m'''}T_k(\rho)_{m'''m}}{\sum_{m'} \mu_{ki}'^*(t_0)_{mm'}\mu_{ik}'(t_0)_{m'm}}.$$

Hence the correlation function φ_{ki} is given by

$$\varphi_{ki}(\tau) = \sum_{m'} \mu_{ki}'^*(t_0)_{mm'}\mu_{ik}'(t_0) \exp(-Nv\sigma\tau),$$

where σ denotes the complex cross section defined by the integral in Eq. (84). It is obvious from the discussion in the previous section that the line shift in frequency and the half-width follow from the imaginary and the real parts of $Nv(1-\sigma)$, respectively.

The cross section σ is given by the average of the partial cross section $\sigma(J_2)$ due to the different state J_2 of perturbers labeled 2.

$$\sigma = \sum_{J_2} B(J_2)\sigma(J_2). \qquad (85)$$

$B(J_2)$ is the Boltzmann distribution function for the colliding molecule in the state J_2. It has been assumed in Eq. (85) that no transitions occurs in molecule (2)

quantum numbers, except perhaps for the degenerate ones, or that transitions in molecules (2) can occur only simultaneously with those of molecule (1), so that T_i and T_k are still essentially diagonal in molecule (1) and (2) (resonance cases). Since the interaction Hamiltonian $C'(t)$ depends on J_2, $\sigma(J_2)$ is averaged over the degenerate substate M_2 (possibly magnetic quantum number) belonging to J_2:

$$S(\rho,J_2)=1-\frac{\sum_{M_2m'm''m'''}\mu_{ki}'^*(t_0)_{mm'}T_i^{-1}(\rho)_{m'm''}{}^{J_2}\mu_{ik}'(t_0)_{m''m'''}T_k(\rho)_{m'''m}{}^{J_2}}{(2J_2+1)\sum_{m'}\mu_{ki}'^*(t_0)_{mm'}\mu_{ik}'(t_0)_{m'm}}. \tag{86}$$

To evaluate this expression S, the expansion (78) may be used. It is found that the zeroth order in $P(\rho)$ of S vanishes. The first-order terms are, after having averaged over the degenerate index m, which will be denoted by m_k and m_i for the absorber in the state k and i, respectively.

$$S_1(\rho,J_2)=i\left[\sum_{m_kM_2}\frac{(k,m_k,J_2,M_2|P|k,m_k,J_2,M_2)}{(2J_k+1)(2J_2+1)}-\sum_{m_2M_2}\frac{(i,m_2,J_2,M_2|P|i,m_i,J_2,M_2)}{(2J_i+1)(2J_2+1)}\right]. \tag{87}$$

Since this is imaginary, S_1 contributes to the line shift. On the other hand, the first term which gives a contribution to the line breadth is one of the two possible second-order terms called by Anderson $(S_2)_{\text{outer}}$.[52]

$$S_2(\rho,J_2)_{\text{outer}}=\tfrac{1}{2}\left[\sum_{m_kM_2}\frac{(k,m_k,J_2,M_2|P^2|k,m_k,J_2,M_2)}{(2J_k+1)(2J_2+1)}+\sum_{m_iM_2}\frac{(i,m_i,J_2,M_2|P^2|i,m_i,J_2,M_2)}{(2J_i+1)(2J_2+1)}\right]. \tag{88}$$

m_k is the degenerate index of an absorber in the state k and $2J_k+1$ is the degeneracy of the state. Another second-order term $S_2(\rho,J_2)_{\text{middle}}$, which arises from the cross term, $\mu_{ki}'^*(t_0)_{mm'}P_i^{-1}(\rho)_{m'm''}{}^{J_2}\mu_{ik}'(t_0)_{m''m'''}P_k(\rho)_{m'''m}{}^{J_2}$ when Eq. (78) is introduced into Eq. (86), is quite complex and difficult to evaluate. But, when the absorber has a permanent electric dipole moment, $S_2(\rho,J_2)_{\text{middle}}$ vanishes. When ρ is small, Eq. (78) cannot be used. In this strong collision, S has its maximum value of $S=1$ (real).

A typical matrix element necessary to evaluate Eqs. (87) and (88) is given by

$$(\alpha|P|\beta)=(k,m_k,J_2,M_2|P|k',m_k',J_2',M_2')$$
$$=\frac{1}{h}\int_{-\infty}^{\infty}\exp(i\omega_{\alpha\beta}t)(\alpha|C(t)|\beta)dt, \tag{89}$$

where

$$\omega_{\alpha\beta}=\frac{E(\alpha)-E(\beta)}{h}=\frac{E(k)+E(J_2)-E(k')-E(J_2')}{h}.$$

$E(k)$ is the energy of the absorbing molecule in the initial state. Clearly from Eq. (89) those $(\alpha|P|\beta)$ where $\omega_{\alpha\beta}=0$ (resonant collision) make a significant contribution to S.

C. STATISTICAL THEORY

1. Classical Statistical Theory

As will be discussed later, many experimental results are inconsistent even with the linear dependence of half-width on density obtained by the phase shift approximation. This discrepancy seems to be more prominent at higher densities, where multiple interactions[30] may dominate binary ones.

To a first approximation molecules can be regarded as at rest, and the frequency emitted by an active molecule at each position x is simply given by $(V_b(x)-V_a(x))/\hbar$, $V_b(x)$ and $V_a(x)$ being potential energies of the molecule concerned in the initial and final states respectively at a point x in configuration space. The intensity within a certain frequency interval $d\omega$ is, therefore, proportional to the probability of those configurations of perturbers which yield the value $V(x)$. This consideration was put into the treatment first by Holtsmark[69] and others[70,53] in connection with the Stark effect in ionized or dipole gases and later by Kuhn[71] and Margenau.[2,72,73]

The potential $V(x)$ of the radiating molecule is the sum of the energy in the isolated state E_0 and perturbation energies due to all other molecules. Thus, it is a function of the position of all perturbing molecules. If we assume additiveness of perturbation energies, we find

$$V(x)=E_0+\sum\hbar\gamma R_i^{-p}U(\xi_i),$$

where R_i is the distance of the ith perturber from the radiating molecule, ξ_i denotes a set of angle and spin variables, and γ is the force constant. As a rule the perturbation energies are much larger for excited states than for normal states. The frequency difference $\Delta\omega$ from the natural frequency will not be given by Eq. (52), but rather will be given approximately by

$$\Delta\omega=\sum_i\gamma R_i^{-p}U(\xi_i), \tag{90}$$

[69] H. Holtsmark, Ann. Physik 58, 577 (1919); Physik. Z. 20, 162 (1919); 25, 73 (1924).
[70] L. Spitzer, Jr., Phys. Rev. 55, 699 (1939).
[71] H. Kuhn, Phil. Mag. 18, 987 (1934); Phys. Rev. 52, 133 (1937).
[72] H. Margenau, Phys. Rev. 40, 387 (1932); 48, 755 (1935).
[73] H. Margenau, Phys. Rev. 82, 156 (1951).

using only γ's for excited states. The variation of intensity with $\Delta\omega$ is determined by the probability $I(\Delta\omega)$ that, in the presence of n perturbing molecules, the potential energy

$$\sum_{i=j}^{n} \gamma R_i^{-p} U(\xi_i)$$

shall have the value $\Delta\omega$. This probability is easily calculated by the use of Markoff's method,[74] or the result may be obtained by the assumption that the probability is proportional to the time interval during which $\Delta\omega$ is radiated. Since perturbers are considered as at rest, the radiation may be expressed by

$$\exp(i \sum_{i=1}^{n} \gamma R_i^{-p} U(\xi_i) t). \tag{91}$$

Then, the Fourier component with respect to t will give the relative intensity as a function of $\Delta\omega$

$$I(\Delta\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(i \sum_{i=1}^{n} \gamma R_i^{-p} U(\xi_i) t - i\Delta\omega t) dt. \tag{92}$$

This has been analytically obtained[75] from Eq. (81) for a limiting case consistent with the above assumption. This must be averaged over the distribution of R_i and ξ_i.

As Margenau[73] assumed, $U(\xi)$ leads very often only to different numerical factors, -1 or $+1$ being equally probable, according to values of ξ. Thus, the probability densities with respect to R_i and ξ_i are given by

$$p(R_i)dR_i = \frac{3R_i^2 dR_i}{R_0^3}, \quad p(\xi_i)d\xi_i = \tfrac{1}{2}d\xi,$$

where $4\pi R_0^3/3$ is the volume occupied by the gas. Then,

$$I(\Delta\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\Delta\omega t} A_n(t) dt, \tag{93}$$

with

$$A_n(t) = \left\{ \int_{-1}^{1} p(\xi) d\xi \int_0^R p(R) dR \exp[i\gamma R^{-p} U(\xi) t] \right\}^n.$$

The evaluation is rather straightforward and the final result is

$$I(\Delta\omega) = \frac{1}{\pi} \int_0^{\infty} \exp(-N g_p t^{3/p}) \cos\Delta\omega t dt, \quad p > \tfrac{3}{2}, \tag{94}$$

where N is the number density of perturbers and

$$g_p = \frac{4\pi}{3} \gamma^{3/p} \sin\left[\left(1 - \frac{3}{p}\right)\frac{\pi}{2}\right] \Gamma\left(1 - \frac{3}{p}\right).$$

It is easily seen that, because of the above assumption as to $U(\xi)$, the intensity distribution given by (93) is symmetrical about the natural frequency. The intensity at the center, where $\Delta\omega = 0$, is

$$I(0) = \left(\frac{p}{3\pi}\right) \Gamma\left(\frac{p}{3}\right) \Big/ (N g_p)^{p/3},$$

which decreases with $N^{-p/3}$ in the normalized $I(\Delta\omega)$. It suggests that, in contrast to Eq. (54a), the half-width increases roughly with $N^{p/3}$. For large values of $\Delta\omega$, $I(\Delta\omega)$ becomes proportional to $\Delta\omega^{-1-3/p}$. Hence, the intensity at the wing of its distribution changes as $\Delta\omega^{-1-3/p}$. If only binary interactions are considered, this also leads to $\Delta\omega^{-1-3/p}$, as can be seen easily by putting $N = 1$ in Eq. (94). This is in agreement with expectation, since the large interaction energies, corresponding to very short interacting distances, are predominantly caused by single impacts of the molecules. According to Holstein's analysis,[58] for large values of $\Delta\omega$, the Fourier integral formula (40) can be transformed to the statistical one. Thus a line shape at the wing is approximated by the statistical theory rather than by the phase shift approximation.

There are other types of forces which do not change sign with ξ, as in the interaction between nonpolar spherically symmetric molecules. Then $U(\xi) = 1$ and the intensity $I(\Delta\omega)$ appears only on one side of the natural frequency. Equation (93) leads to

$$I(\Delta\omega) = \frac{1}{\pi} \int_0^{\infty} \exp(-N g' t^{3/p}) \cdot \cos(\Delta\omega t + N g'' t^{3/p}) dt, \tag{95}$$

where

$$g' + ig'' = i\frac{4\pi}{3} \gamma^{3/p} \int_0^{\infty} \rho^{-3/p} e^{i\rho} d\rho.$$

For van der Waals forces, for which $p = 6$, $\gamma < 0$ and $g' = g''$,

$$I(\Delta\omega) = \begin{cases} 0, & \text{if } \Delta\omega < 0, \\ \dfrac{2\pi}{3} |\gamma|^{\frac{1}{2}} N \Delta\omega^{-\frac{3}{2}} \exp\left(-\dfrac{4\pi}{9} |\gamma| N^2/\Delta\omega\right), & \text{if } \Delta\omega > 0, \end{cases} \tag{96}$$

as was derived by Margenau,[2,72] and the shift and half-width changes as

$$\Delta\omega_m = -(\tfrac{2}{3}\pi)^3 |\gamma| N^2, \quad \text{(red shift)} \tag{97}$$

$$\Delta\omega_{\frac{1}{2}} = 0.822\pi^3 |\gamma| N^2. \tag{98}$$

This yields a method for obtaining the van der Waals coefficient for excited states. Since Eq. (96) involves the assumption that a force varying as R^{-6} is valid all the way into the origin of R, Eq. (96) may be in error for large $\Delta\omega$, which corresponds to close encounters. But for large $\Delta\omega$, Eq. (96) gives

$$I(\Delta\omega) \sim \Delta\omega^{-\frac{3}{2}}. \tag{99}$$

[74] S. Chandrasekhar, Revs. Modern Phys. 15, 1 (1943).
[75] P. W. Anderson, Phys. Rev. 86, 809 (1952).

This is Kuhn's relation for the wing of the broadened line, as supported by his observation[76] of the broadened wing of Hg 2357 by A and Hg, and by the observations of Ruhmkorf[77] using N_2, O_2, CO_2, and A as perturbers and of Minkowski[77] with NaD lines perturbed by A. This was also confirmed with D lines in an acetylene flame by Huldt and Knall.[12] However, an experimental line shape at the wing seems sometimes to be more complicated.[78,79]

The validity of Eq. (99) has been discussed by Lindholm,[57] Holstein,[58] and Anderson.[75] The latter two obtained for the wing,

$$I(\omega)=\frac{4\pi N\gamma^{3/p}}{p\Delta\omega^{1+3/p}}\times\left\{1-\frac{p}{36}\left(1+\frac{1}{2p}\right)\left(1-\frac{1}{p^2}\right)\frac{v^2}{\Delta\omega^{2-2/p}\gamma^{2/p}}\right\},\qquad(100)$$

for the interaction energy of the type $\Delta\omega=\gamma/R^p$. Thus Eq. (99) is valid when $\Delta\omega\gg v^{p/(p-1)}\gamma^{-1/(p-1)}$. On the other hand, the impact theory is good[59] when $\Delta\omega\ll v^{p/(p-1)}\gamma^{-1/(p-1)}$ (near the unperturbed line). This condition for the validity of the statistical theory may be rewritten, as obtained rather rigorously by Margenau,[80] as $T\Delta\omega\gg 1$, where T is about equal to ρ/v, ρ being the impact parameter for the collision, since $\Delta\omega_{\frac{1}{2}}$ (statistical)$\simeq\gamma N^{p/3}\sim\gamma/\rho^p$.

The recent experimental results[81–86] at pressures of several hundreds of atmospheres reveal the insufficiency of the above theory. Equation (90) does not predict the reversal of the direction of line shift at high pressure which is found in some cases (see Fig. 3). Bergeon, Robin, and Vodar[87] presented a qualitative discussion suggesting that the repulsive force, which would be dominant at high pressures, must be included in the calculation of Margenau's formula (93), in addition to the long-range force given by Eq. (90). They employed the Lennard-Jones formulas for the interaction energy and calculated the line shape numerically, assuming the additiveness of the energy and its applicability to both normal and excited energy levels. Though better broadened line contours than otherwise were obtained, the forces at small distances are not clear without detailed discussion. In any case, it is likely that they are not additive.[88] However, potential curves at very close distance have been devised[89] to explain the absorption bands (see III.C) appearing on the violet side of spectral lines.

Furthermore, recent observations on certain steller absorption line broadening[90] and on shock phenomena[91] have made clear some insufficiencies of the statistical theory of Holtsmark[69] which attributes the broadening to the Stark effect. The Debye-Hückel effect was introduced to account for the discrepancies, but this resulted in corrections which are too small.[53,92] Following the suggestion of Baranger,[93] an extensive study has been made quantum-mechanically on electron impact line broadening, showing that for the Lyman α line, where there is a linear Stark effect, the ion broadening is the dominant effect,[94] but that for the anomalous $3d\to 2p$ line of HeI, which has a small Stark shift, the electrons can make important contributions.[95]

Generally speaking, pressure broadening changes from collision type to statistical type with increasing pressure.[58,59,71] However, as shown in Part III, the transition from one type to the other is not clear. The "modulation index," which corresponds to the ratio of the maximum frequency shift to the number of collisions per second, has been proposed[96] to determine the predominance of one type or the other under certain conditions. Lindholm,[57] Holstein,[97] and others worked out the transition, as detailed in II.D, including the line shape at the violet wing.

2. Quantum-Mechanical Justification of Statistical Theory

Jablonski[98] has given a theory of the pressure-broadened atomic line shape with a very close analogy to the theory of intensity distribution in molecular spectra. He calculated the stationary states of interatomic motion under the influence of an interaction between a radiating atom and a colliding one.

For simplicity, consider the wave function φ of a single pair of atoms consisting of a radiator and a perturber in a spherical box of radius R_0, with the radiator at its center and with the boundary condition $\varphi(R_0)=0$. In the Lorentz approximation, where molecular rotation

---

[76] H. Kuhn, Proc. Roy. Soc. (London) **A158**, 212 (1937); **158**, 230 (1937).
[77] H. A. Ruhmkorf, Ann. Physik **31**, 21 (1938); R. Minkowski, Z. Physik **93**, 731 (1935).
[78] S. Y. Ch'en and C. S. Chang, Phys. Rev. **75**, 81 (1949).
[79] Ch'en, Bennett, and Jefimenko. J. Opt. Soc. Am. **45**, 182 (1956).
[80] H. Margenau, RM-1670-AEC, April 3 (1956).
[81] S. Y. Ch'en, thesis, California Institute of Technology (1940), p. 35.
[82] Robin, Robin, and Vodar, Compt. rend. **232**, 1754 (1951).
[83] S. Robin, thesis, Paris, (1951); J. chim. Phys. **49**, 1 (1952).
[84] J. Robin and F. Robin, Compt. rend. **233**, 928 (1951).
[85] J. Robin and S. Robin, Compt. rend. **233**, 1019 (1951); J. Robin, thesis, Paris (1953).
[86] (a) J. Robin, Compt. rend. **238**, 1491 (1954); (b) A. Michel and H. De Kluiver, Physica **22**, 919 (1956).
[87] Bergeon, Robin, and Vodar, Compt. rend. **235**, 360 (1952); R. Bergeon, *ibid.* **238**, 2507 (1954); R. Bergeon and B. Vodar, *ibid.* **240**, 172 (1955).
[88] L. Jansen and Z. I. Slawsky, J. Chem. Phys. **22**, 1701 (1954); R. T. McGinnies and L. Jansen, Phys. Rev. **101**, 1301 (1956).
[89] H. Kuhn, Z. Physik **72**, 462 (1931); and see reference 146, 147, and 149.
[90] A. B. Underhill, Astrophys. J. **116**, 446 (1953); G. L. Odgers, Astrophys. J. **116**, 444 (1953).
[91] A. Kantrowitz, Phys. Rev. **90**, 368 (1953).
[92] F. N. Edmonds, Astrophys. J. **123**, 45 (1956).
[93] M. Baranger, Phys. Rev. **91**, 436 (1953).
[94] Kivel, Bloom, and Margenau, Phys. Rev. **98**, 495 (1955); H. Margenau and B. Kivel, Phys. Rev. **98**, 1822 (1955).
[95] B. Kivel, Phys. Rev. **98**, 1055 (1955).
[96] H. Margenau and S. Bloom, Phys. Rev. **79**, 213 (1950).
[97] G. N. Plass and D. Warner, Phys. Rev. **86**, 138 (1952).
[98] A. Jablonski, Acta Phys. Polon. **6**, 371 (1937); **7**, 196 (1938); Physica **8**, 541 (1940); Phys. Rev. **68**, 78 (1945).

can be neglected, the relevant part of φ is only the radial wave function φ_{vl}, where v and l are a set of radial and angular momentum quantum numbers for the relative motion of the two atoms. Since the intermolecular forces are different for upper and lower states, which are designated by $'$ and $''$, respectively, the solutions of two different radial wave equations:

$$d\varphi_{v'l'}/dR+\frac{2m}{\hbar^2}\left(W_{v'l'}-V'(R)-\hbar^2\frac{l'(l'+1)}{2mR^2}\right)\varphi_{v'l'}=0, \quad (101)$$

and

$$d\varphi_{v''l''}/dR+\frac{2m}{\hbar^2}\left(W_{v''l''}-V''(R)-\hbar^2\frac{l''(l''+1)}{2mR^2}\right)\varphi_{v''l''}=0,$$

are required. m denotes the reduced mass of the pair, W_{vl} the relative energy of the motion, and $V(R)$ the mutual potential energy.

It is easily shown that the two sets of rotational wave functions for upper and lower states are orthogonal. Hence, only transitions with $l'=l''$ are allowed. If we denote the atomic transition amplitude by $\mathfrak{M}_0$, we may write the total transition amplitude as

$$D_{v'v''l}=\mathfrak{M}_0\int\varphi_{v'l}{}^*\varphi_{v''l}dv. \quad (102)$$

Here we have assumed the classical Franck-Condon principle.[99] Clearly the matrix element $D_{v'v''l}$ depends on l through the solution φ_{vl} of (101). For large l's in comparison with the difference of $V'(R)$ and $V''(R)$, both sets $\varphi_{v'l}$ and $\varphi_{v''l}$ become practically identical and hence orthogonal. Therefore, no transitions $v'\rightarrow v''$ with $v'\neq v''$ will occur between states of sufficiently large l's, and there will be no appreciable change of translational energy. Thus, this region of large l's corresponds to Lorentz' impact parameters larger than the critical optical collision diameter. However, for smaller values of l, the dependence of $D_{v'v''l}$ on l becomes very marked. Classically speaking, the effect on spectral lines produced by collisions depends very markedly on impact parameter. Thus, one is not justified in assuming that all collisions with impact parameter smaller than a certain optical collision diameter are equally effective.[1,54,55]

The solution of Eq. (101) with WKB approximation is

$$\varphi_{vl}=\left(\frac{2p(\infty)}{R_0p(R)}\right)^{\frac{1}{2}}\cos\left[\frac{1}{\hbar}\int_{R_t}^{R}p(x)dx+\delta\right], \quad (103)$$

where $p(x)=[2m(W_{vl}-V(x)-\hbar^2l(l+1)/x^2)]^{\frac{1}{2}}$ is the radial component of relative momentum of the two atoms. $p(\infty)$, which is the value of $p(x)$ at $x=\infty$, together with R_0 in Eq. (103), is determined by normalization. The above solution is valid only for $R>R_t$, where R_t is the distance of classical turning point given by $p(R_t)=0$. δ is a phase constant which is adjusted according to the solution for $R<R_t$. The transition amplitude is

$$D_{v'v''l}=\frac{\mathfrak{M}_0}{R_0}\int_{R_t}^{R_0}\left(\frac{p'(\infty)p''(\infty)}{p'(R)p''(R)}\right)^{\frac{1}{2}}\times\cos\left(\int_{R_t}^{R}\frac{p'(x)-p''(x)}{\hbar}dx+\delta'-\delta''\right)dR. \quad (104)$$

Obviously, the largest contribution to the value of this integral comes from the region $R\sim R_c$ for which $p'(R)=p''(R)$. Thus, the relative momentum $p(R)$ of the nuclei of the two atoms does not change in the transition $v'\leftrightarrow v''$, consistent with the classical conservation law of momentum. It may be interpreted, therefore, that the transition $v'\leftrightarrow v''$ occurs at $R=R_c$. Then, the energy change of nuclear motion involved in the transition is given by the condition $p'(R_c)=p''(R_c)$, i.e.,

$$W_{v'l}-W_{v''l}=V'(R_c)-V''(R_c). \quad (105)$$

This energy must be transferred to the radiation, resulting in the frequency shift $\Delta\omega$. Thus, $\Delta\omega$ may be expressed as a function of R_c. Then, if $D_{v'v''l}$ of Eq. (104) is known as a function of R_c, the probability of the frequency shift $\Delta\omega$ will result.

To obtain the function $D_{v'v''l}(\Delta\omega)$, Jablonski first expanded the integrand of Eq. (104) near $R=R_c$ and took only the first nonvanishing two terms. This procedure depends primarily on the assumption that the integrand of Eq. (104) slowly oscillates near $R=R_c$. This assumption is clearly equivalent to the assumption made by Kuhn and London.[55]

In the calculation of $I(\Delta\omega)$ for absorption of a particular atomic line, the probability $D^2_{v'v''l}$ must be averaged over all initial and final states, $v''l$ and $v'l$. Thus, after some reductions, the asymptotic form is

$$I(\Delta\omega)=\frac{3R_c{}^2(1-V''(R_c)/W_{v''l})^{\frac{1}{2}}}{R_0{}^3|\hbar d\Delta\omega/dR|_{R=R_c}}, \quad (106)$$

where $\hbar\Delta\omega=V'(R_c)-V''(R_c)$. This can be rewritten as

$$I(\Delta\omega)d\omega=CR_c{}^2dR_c,$$

where C is some constant. This is clearly an expression of Kuhn's idea.[71] The relation, like Eq. (106), for emission is obtained by replacing $''$ by $'$.

If n free perturbers exist in the box and the interaction is additive, the total wave function $\varphi_{vl}{}^t$ and the total energy will be such that

$$\varphi_{vl}{}^t=\prod_i^n\varphi_{vl,i},\quad W_{vl}{}^t=\sum_i^n W_{vl,i},$$

[99] G. Herzberg, *Molecular Spectra and Molecular Structure* (D. Van Nostand Company, Inc., New York, 1951).

with

$$V'(R_c)^t - V''(R_c)^t = \hbar\Delta\omega. \tag{107}$$

Again, ′ designates the upper state. Then, since $n=(4\pi/3)R_0^3N$, where N is the number density of perturbers, and since the energy is additive,

$$I^t(\Delta\omega) = \sum_i a_i 4\pi N f^2(\Delta\omega)\frac{(1-V''(R_c)/W_{v''l})}{|\hbar d\Delta\omega/dR_c|} \tag{108}$$

where a_i assures the different probability arising from the different type of potential energy $V(R)$, if any. $f(\Delta\omega)$ is R_c expressed as a function of $\Delta\omega$ obtained from Eq. (107). Strictly speaking, $I^t(\Delta\omega)$ must be averaged over all W_{vl} occurring in the gas under consideration, but, since the factor in brackets constitutes a small correction only, at least in a certain range of frequency, it is permissible to put $W_{vl}=\frac{3}{2}kT$. Thus, the dependence on temperature of $I^t(\Delta\omega)$ is very slight, as experimentally observed, except for the part of the line arising from transitions in the repulsive branches of the potential curves, particularly at the points at which $V(R_c)=W_{vl}$. Apart from the above factor, Eq. (108) is identical with the equation obtained from the statistical theory.[2,72]

Since the WKB method is valid when the potential energy $V(R)$ changes so slowly that the momentum of the particle is sensibly constant over many wavelengths of φ_{vl}, Eq. (108) cannot be used for light perturbers or light radiators such as helium.

D. THEORY AT GENERAL PRESSURES

There have been many attempts[2,57,58,75,96,97] to span a bridge between the impact theory and the statistical theory, especially to find a theory valid at intermediate pressures, so that the broadening behavior, including the line shape, over the whole range of pressures is described. In practice, however, more stress was put on the line shape at the wings with conditions such as in Eq. (100) because of the mathematical difficulties involved.

Recently Anderson and Talman[100] extended the use[75] of the general Fourier integral formula, (40), in a more general way. Equation (40) was obtained under the assumption of classical paths of perturbers and the adiabatic assumption. It goes over to the expression in either the impact theory or the statistical theory depending on the additional assumptions made simply for an easier mathematical treatment. One of these made here is about the frequency perturbation: $\Delta\omega=-(\gamma/R^p)$ as in Eq. (90). Since this interaction energy is usually additive and if we assume that the motions of perturbers are independent, the correlation function (45) can be rewritten:

$$\varphi(\tau)=\left\langle \exp\left[-i\int_0^\tau \Delta\omega(R)dt'\right]\right\rangle^n. \tag{109}$$

n is the total number of perturbers. As in Eq. (46), the average may be understood as over the configuration space. Since collisions are isotropic and independent, we assume without loss of generality that all collisions take place from the direction of x. Referring to generalized coordinates ρ, x_0, the impact parameter and the position on the trajectory at $t'=0$ respectively, one finds that the volume, V, of our system is represented by

$$V=2\pi\int\rho d\rho\int dx_0. \tag{110}$$

If we assume, in addition, that the perturbers describe straight line paths and all travel with velocity $\bar{v}$; $R^2=(x_0+\bar{v}t')^2+\rho^2$. Thus,

$$\varphi(\tau)=\left\{\frac{1}{V}\int dV\exp\left[-i\int_0^\tau \Delta\omega(R)dt'\right]\right\}^n,$$

$$=\left\{1-\frac{2\pi}{V}\int_0^\infty\rho d\rho\int_{-\infty}^\infty dx_0\right.$$

$$\left.\times\left[1-\exp\left(-i\int_0^\tau\Delta\omega\{[(x_0+\bar{v}t')^2+\rho^2]^{\frac{1}{2}}\}dt'\right)\right]\right\}^n,$$

$$\equiv\left[1-\frac{V'(\tau)}{V}\right]^n,\text{ say.} \tag{111}$$

Then, by making V and $n\to\infty$, while the number density of perturber $N=n/V$ is kept constant, we have

$$\varphi(\tau)=\exp[-NV'(\tau)]. \tag{112}$$

This is the basic formula.

For small values of $\bar{v}$, Eq. (112) yields the statistical distribution which is exact if the perturbers are stationary. With $\bar{v}=0$ in Eq. (112), we obtain

$$V'(\tau)=4\pi\int_0^\infty R^2\{1-\exp[-i\Delta\omega(R)\tau]\}dR, \tag{113}$$

noticing that $\rho^2+x_0^2=R^2$ and $2\pi\rho d\rho dx_0=4\pi R^2dR$. A simple calculation will show that Eq. (113) gives Margenau's equation (93). Equation (113) is valid not only as $v\to0$ but as $\tau\to0$. If one refers to Eq. (112), it becomes clear that when N is large $\varphi(\tau)$ is only large for small τ so that the above is indeed also the high pressure limit.

For large $\bar{v}$ and low density Eq. (112) should approach the impact theory. Putting $u=x_0+\bar{v}\tau$, one finds that $V'(\tau)$ becomes

$$V'(\tau)=2\pi\int_0^\infty\rho d\rho\int_{-\infty}^\infty dx_0$$

$$\times\left\{1-\exp\left[-i\int_{x_0}^{x_0+\bar{v}\tau}\Delta\omega((\rho^2+u^2)^{\frac{1}{2}})\frac{1}{\bar{v}}du\right]\right\}.$$

[100] P. W. Anderson and J. D. Talman, Bell Telephone Laboratories, Murray Hill, New Jersey (1956).

The integrand in the x_0 integration is small whenever the exponent

$$\frac{1}{\bar{v}}\int_{x_0}^{x_0+\bar{v}\tau}\Delta\omega((\rho^2+u^2)^{\frac{1}{2}})du$$

is small, that is, when $x_0 \ll -\bar{v}\tau$ or $x_0 \gg 0$, since the appreciable values of this integral lie in the range of $-\bar{v}\tau < x_0 < 0$.|| Hence,

$$V'(\tau) \sim 2\pi\int_0^\infty \rho d\rho \int_{-\bar{v}\tau}^0 dx_0$$
$$\times\left\{1-\exp\left[-i\int_{-\infty}^{\infty}\Delta\omega((\rho^2+u^2)^{\frac{1}{2}})\frac{1}{\bar{v}}du\right]\right\}, \quad (114)$$
$$=2\pi\bar{v}\tau\int_0^\infty \rho d\rho(1-e^{-i}\Theta(\rho)),$$

with

$$\Theta(\rho)=\int_{-\infty}^{\infty}\Delta\omega((\rho^2+u^2)^{\frac{1}{2}})\frac{1}{\bar{v}}du.$$

Referring to Eq. (53), we note that $\Theta(\rho)$ is a phase shift at a collision. It is clear now that Eq. (114), with Eq. (112), gives Foley's expression (50).

Equations (113) and (114) are the two limits of V' for large and small t. To study the intermediate region we choose a dimensionless form. For the frequency perturbation $\Delta\omega=-\gamma/R^p$, we define $R_0=(\gamma/\bar{v})^{1/(p-1)}$. With new variables such that $\rho=R_0r$, $x=R_0z$, $N=h/R_0^3$, $\Delta\omega=\bar{v}\Delta\omega'/R_0$, $t=R_0u/\bar{v}$ and $\tau=R_0y/\bar{v}$, Eq. (112) becomes

$$\varphi(R_0y/\bar{v})=e^{-h\psi(y)}, \quad (115)$$

with

$$\psi(y)=2\pi\int_0^\infty rdr\int_{\infty-}^{\infty}dz(1-e^{i\Theta}), \quad (116)$$

$$\Theta(y)=\int_0^y[(z+u)^2+r^2]^{-p/2}du. \quad (117)$$

Hence, from Eq. (44) or (81), we have

$$I(\bar{v}\Delta\omega'/R_0)=\frac{R_0}{2\pi\bar{v}}\int_{-\infty}^{\infty}\exp[i\Delta\omega'y-h\psi(y)]dy, \quad (118)$$

neglecting less important factors. $I(\Delta\omega')$ is determined by one parameter only, $h=NR_0^3=N(\gamma/\bar{v})^{3/(p-1)}$, and hence it is the magnitude of this parameter which determines the validity of the statistical or impact theories. Lenz's[54] method is equivalent to the approximation in Eq. (117) of replacing $\Theta(y)$ by its tangent at $y=0$ and is in disagreement with the statistical theory. Here notice that $y=\bar{v}\tau/R_0$. Lindholm's[57] method, which was carried out for a perturbation proportional to $1/R^6$ with a simplified model of collisions during which the frequency was constant, is equivalent to the approximation of replacing $\Theta(y)$ by a straight line which cuts the curve $\Theta(y)$ at $y=0$ and large. Thus, Lindholm obtained a result in agreement also with the statistical theory in its extremities, which could, however, be wrong in intermediate cases.

The problem of calculating a line shape has been reduced to that of calculating the dimensionless function (116). However, $\psi(y)$ cannot be obtained explicitly and is difficult to compute numerically. Anderson and Talman used the method to calculate the behavior of $\psi(y)$ for very small values of y and for large values of y and then to interpolate between the two.

They gave, for the case $p=6$, and small y

$$\psi(y)=i^{-1/2}\frac{4\pi}{3}\Gamma(\tfrac{1}{2})y^{1/2}+i^{-1/6}\frac{5\pi}{18}\Gamma(\tfrac{5}{6})y^{13/6}, \quad (119)$$

and for large y

$$\psi(y)=\frac{2\pi}{5}i^{-2/5}\left(\frac{3\pi}{8}\right)^{-3/5}\Gamma(\tfrac{2}{5})y+1.91+2.63i. \quad (120)$$

The asymptotic behaviors of the two match very well so that these may be used to approximate $\psi(y)$. Equations (119) and (120) will give numerically a line shape at any pressure for $p=6$, but no general analytic expression for the line shape can be obtained.

We shall be interested in the asymptotic behavior of $I(\omega)$ for: (a) large positive $\Delta\omega$, (b) $\Delta\omega$ near zero, and (c) large negative $\Delta\omega$, when h is large. Cases (a) and (b) cannot be properly treated by the statistical theory, as discussed before. For large values of h, Eq. (118) is large only for $|y|$ small, since the real part of $\psi(y)$ is a monotonically increasing function of $|y|$. Hence, $\psi(y)$ may be replaced by its expansion near the origin.

$$\psi(y)=\frac{4\pi}{3}\Gamma\left(1-\frac{3}{p}\right)\left(\frac{y}{i}\right)^{3/p}$$
$$-\frac{(p-1)\pi}{18}\Gamma\left(1-\frac{1}{p}\right)\left(\frac{y}{i}\right)^{2+(1/p)}. \quad (121)$$

This is a general expression for Eq. (119). Therefore, we have

$$I(\bar{v}\Delta\omega'/R_0)=\frac{R_0}{2\pi\bar{v}}\int_{-\infty}^{\infty}\exp[\phi(\Delta\omega',y)]dy, \quad (122)$$

where

$$\phi(\Delta\omega',y)=i\Delta\omega'y-h\frac{4\pi}{3}\Gamma\left(1-\frac{3}{p}\right)\left(\frac{y}{i}\right)^{3/p}$$
$$+h\frac{(p-1)\pi}{18}\Gamma\left(1-\frac{1}{p}\right)\left(\frac{y}{i}\right)^{2+(1/p)}. \quad (123)$$

In case (a), $\Delta\omega' \gg h$. The function $\exp(i\Delta\omega'y)$ is a rapidly oscillating function and e^ϕ oscillates rapidly except at points y_0 such that $d\psi/dy=i\Delta\omega$. Namely, the

|| This corresponds to the physical interpretation that only those which pass by the absorber within the time $0<t'<\tau$ affect the line shape.

neighborhoods of the points satisfying the equation,

$$\phi'(\Delta\omega',y_0)=i\Delta\omega'-\frac{4\pi}{p}\frac{h}{i}\Gamma\left(1-\frac{3}{p}\right)\left(\frac{y_0}{i}\right)^{(3/p)-1}$$

$$+\left(2+\frac{1}{p}\right)\frac{h}{i}\frac{(p-1)\pi}{18}\Gamma\left(1-\frac{1}{p}\right)\left(\frac{y_0}{i}\right)^{1+(1/p)}=0, \quad (124)$$

can contribute appreciably to the integral (122). When $p>3$, this is satisfied by y/i a large positive number, the third term being appreciable, or by y/i a small negative number, the second term being appreciable. From the branch cut of the integral (122), the first choice must be made and

$$y_0\sim i\left[\frac{18p\Delta\omega'}{h(p-1)(2p+1)\pi}\Big/\Gamma\left(1-\frac{1}{p}\right)\right]^{p/(p+1)}. \quad (125)$$

To integrate Eq. (122) around this point, ϕ may be expanded into Taylor series

$$\phi=\phi(\Delta\omega',y_0)+\tfrac{1}{2}\phi''(\Delta\omega',y_0)(y-y_0)^2+\cdots.$$

Thus, we have

$$I(\bar{v}\Delta\omega'/R_0)\sim\frac{R_0}{2\pi\bar{v}}\exp[\phi(\Delta\omega',y_0)]$$

$$\times\int_{-\infty}^{\infty}\exp[\tfrac{1}{2}\phi''(\Delta\omega,y_0)(y-y_0)^2]dy$$

$$=\frac{R_0}{\bar{v}}(-2\pi\phi'')^{-\frac{1}{2}}e^{\phi}. \quad (126)$$

This, with Eq. (125), gives the asymptotic behavior of $I(\Delta\omega')$. However, for $\Delta\omega'\gg h$, the intensity will be so low that it cannot be observed.

In case (b) ($\Delta\omega'$ small), on the other hand, it is of greater interest. The same method[101] as above will be applied. Since $\Delta\omega'\sim 0$, we have, from Eq. (124),

$$y_0=i\left\{\frac{72\Gamma[1-(3/p)]}{(p-1)(2p+1)\Gamma[1-(1/p)]}\right\}^{p/2(p-1)}. \quad (127)$$

Hence, Eq. (126) gives

$$I(\bar{v}\Delta\omega'/R_0)=\left(\frac{1}{16\pi^2h\Gamma[1-(3/p)]}\frac{p^2}{p-1}\right)^{\frac{1}{2}}$$

$$\times\left(\frac{(p-1)(2p+1)\Gamma[1-(1/p)]}{72\Gamma[1-(3/p)]}\right)^{(3-2p/4(p-1))}$$

$$\times\exp\left[-2h\left(\frac{p-1}{2p+1}\right)\frac{4\pi}{3}\Gamma\left(1-\frac{3}{p}\right)\left(\frac{72\Gamma[1-(3/p)]}{(2p+1)(p-1)\Gamma[1-(1/p)]}\right)^{3/(2p-2)}\right] \quad (128)$$

$$\times\exp\left[-\Delta\omega'\left(\frac{72\Gamma[1-(3/p)]}{(p-1)(2p+1)\Gamma[1-(1/p)]}\right)^{p/(2p-2)}\right].$$

Thus, at the position of the unperturbed frequency, the intensity falls off exponentially with both increasing h and increasing $\Delta\omega$. Both of these results disagree with Lindholm's result[57] such as const.$\times\Delta\omega^{-7/3}$ on this (violet) wing. However, the difference is very small[97] within the range of validity of the statistical theory. Actually, Huldt and Knall[12] observed that the intensity at the violet wing falls off as $\Delta\omega^{-2.35}$ which is very close to Lindholm's result.

For the red wing, $\Delta\omega'\ll 0$, Anderson and Talman obtained Eq. (100), the same result as obtained by Holstein, for large h. They showed that it is also valid for all h.

For $\Delta\omega'$ and h both small, the integrand in Eq. (118) is independent of $\Delta\omega'$ for $y<1$ since $\exp i\Delta\omega' y$ varies very little in this range. Therefore, the line shape should come from large y, say, Eq. (120) for $p=6$.

$$I(\bar{v}\Delta\omega'/R_0)=\frac{R_0}{2\pi\bar{v}}\int_{-\infty}^{\infty}\exp(i\Delta\omega' y-h\sigma y-hC)dy, \quad (129)$$

where

$$\sigma=\sigma_0 i^{-2/5},\quad \sigma_0=\frac{2\pi}{5}\left(\frac{3\pi}{8}\right)^{-3/5}\Gamma(\tfrac{2}{5}),$$

$$C=1.91+2.63i.$$

Thus, we have

$$I(\bar{v}\Delta\omega'/R_0)=e^{-1.91h}\left[\frac{\sigma_0 h\cos(\pi/5)\cos 2.63h-\sin 2.63h[\Delta\omega'-\sigma_0 h\sin(\pi/5)]}{[\Delta\omega'-\sigma_0 h\sin(\pi/5)]^2+h^2\sigma_0^2\cos^2(\pi/5)}\right]. \quad (130)$$

[101] This method is called the method of steepest descents. See G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, New York, 1944), p. 235, or P. M. Morse and H. Feshbach, *Method of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 440.

Therefore, the impact theory is valid only if the asymmetry term is small, namely, if $\Delta\omega'-\sigma_0 h\sin(\pi/5)$ is small for small h, as obtained by Spitzer.[54] Hence, the impact theory is applicable in the region near the peak. However, the center of the line at moderate pressures must be discussed by a numerical integration of Eq. (118) using Eqs. (119) and (120) for $\psi(y)$. Because of the fact that the line shapes thus calculated are almost indistinguishable from Lindholm's lines, for the central portion of the line the small y may reasonably correspond to the statistical theory and the large y to the impact theory.

It may be noticed that the above theory is exact when its basic assumptions about intermolecular forces are correct. Thus, in view of the fact that the intermolecular forces assumed above are not entirely realistic, although they have been used widely in theories, recalculation must be made. In this sense, the results[58] might be much improved, especially on the violet wing, if Eq. (118) is numerically integrated for the Lennard-Jones potential.

E. FINE STRUCTURE PRESSURE BROADENING AT ATOMIC LINES

Many experimental results have shown that the pressure-broadening scheme is different for different multiplet components. Half-widths, shifts, and asymmetries are definitely dependent on J, the inner quantum number.

1. Identical Atoms (See III B.a)

If Furssow and Wlassow's view[23] on resonance broadening (II. A. 1.a) is followed, the problem is reduced to the calculation of the dependence on J of the probability of energy transfer per collision between excited and unexcited identical atoms. This transfer would work as a radiation damping of the radiating atom and gives a broadened line of the dispersion type. Following Foley's calculation,[22] and making use of Eq. (55), we find the result for half-width to be

$$\Delta\omega_{\frac{1}{2}}=4\sqrt{3}\pi^4 N\frac{\hbar e^2}{m_e\Delta W_{JJ'}}\left(\frac{2J+1}{2J'+1}\right)^{\frac{1}{2}} f_{JJ'}, \qquad (131)$$

with

$$f_{JJ'}=\frac{2m_e}{3\hbar^2}\frac{\Delta W_{JJ'}}{2J+1}\sum_{m'}|(J',m'|x|J,m)|^2.$$

Here $\Delta W_{JJ'}$ is the difference of the energies in the ground and excited states with inner quantum numbers J and J', respectively. $(J',m'|x|J,m)$ is the matrix element for dipole radiation, the square of which must be summed over all magnetic quantum numbers m' at the excited state for absorption. N is the number density of perturbers. For alkali doublets, the ratio of the half-width of the $J'=\frac{3}{2}$ and $J'=\frac{1}{2}$ component is approximately $\sqrt{2}$, when we neglect the difference of $\Delta W_{JJ'}$ for the two. Lindholm[102] calculated also quantum-mechanically the widths of Rb doublet components with the result of the ratio 1.6.

2. Perturbation Due to Foreign Atoms[28] (See III B.b)

If we assume the perturbers are spherically symmetric, we find the dispersion force is different for alkali atoms in the states $J=\frac{3}{2}$ and $\frac{1}{2}$ to the extent that the levels for $J=\frac{3}{2}$ split, according as $|m|=\frac{3}{2}$ and $\frac{1}{2}$. Therefore, the half-width for the ${}^2P_{\frac{3}{2}}$ component of the alkali resonance doublet is larger than that for the ${}^2P_{\frac{1}{2}}$ component. Using the value of dispersion energy given in (II. A. 1.b) and Foley's relation (58) with $p=6$, we find the half-width for the ${}^2P_{\frac{3}{2}}$ component is roughly 1.12 times larger than that of the ${}^2P_{\frac{1}{2}}$ component. But since the van der Waals force averaged over m for the two states is the same, no appreciable difference in shift is predicted. The difference of relative shift may come from the repulsive part of the interaction energy. Since the energy separation of the states ${}^2P_{\frac{3}{2}}{}^{|m|=\frac{3}{2}}$ and ${}^2P_{\frac{3}{2}}{}^{|m|=\frac{1}{2}}$ is dependent on the separation of the two atoms, greater asymmetry results for the ${}^2P_{\frac{3}{2}}$ component.

If the foreign gas pressure is high, perturbers may be arranged spherically symmetrically around the radiating alkali atom. Then the quadrupole moment of the atom in $J=\frac{3}{2}$ has no effect on the van der Waals force. Thus, the difference in half-width of the doublet components decreases as pressure increases, although at the wing of the line contour the above effect may still take place.

However, at high pressure, the repulsive part of the interaction energy may be predominant. Then, the energy may be found from the expression (5). The wave function is

$$\varphi=\varphi_n+\sum_j{}'\frac{(j|V|n)}{W_n-W_j}\varphi_j. \qquad (132)$$

Starting from this wave function, one calculates the spin orbit energy, neglecting the atomic core effect,

$$\begin{cases}\frac{1}{2}l\zeta_{nl} & \text{if}\quad j=l+\frac{1}{2},\\ -\frac{1}{2}(l+1)\zeta_{nl} & \text{if}\quad j=l-\frac{1}{2},\end{cases}$$

$$\zeta_{nl}=d_{nl,nl}-(\hbar^2/|\Delta W|m_e^2c^2)\int_{r_0}^{\infty}|R_{nl}|^2\frac{e^2}{r}dr, \qquad (133)$$

where $d_{nl,nl}$ is the unperturbed ζ_{nl} and r_0 is given by the pressure equation $p=-dE/4\pi r_0^2dr_0$ with E from Eq. (5). The separation of the doublet decreases at higher pressure and one concludes that the shift of the ${}^2P_{\frac{3}{2}}$ component is smaller than that of the ${}^2P_{\frac{1}{2}}$, with no appreciable difference in half-width and asymmetry. At very high pressure, the relative magnitudes of the shifts would be reversed because of the distortion of the atomic core and its smaller screening effect on the nuclear charge.[28]

[102] E. Lindholm, Naturwissenschaften 30, 533 (1942).

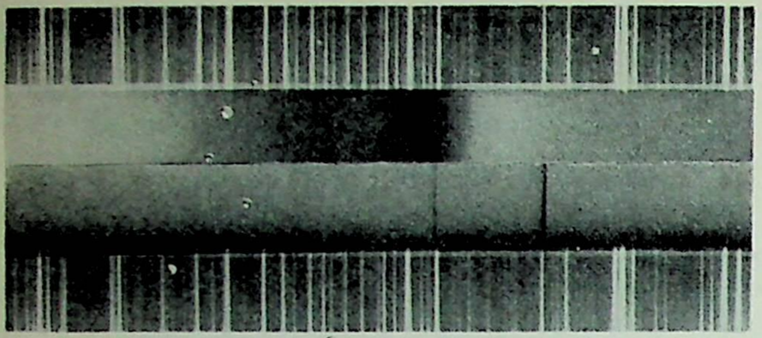


FIG. 1. The red shift and broadening of the blue doublet of Cs produced by argon (after Ch'en and Parker[103]).

Upper spectrum: Argon pressure = 128 atmospheres
Temperature = 262°C
Lower spectrum: Argon pressure = 1.09 atmospheres
Temperature = 150°C.

III. EXPERIMENTAL OBSERVATIONS

A. PRESSURE SHIFT AND BROADENING OF ATOMIC LINES

A review of the essentials of the experimental obser-varions published during the past twenty years on the effects of foreign gases at high pressures on atomic lines is given in the following five subsections. So far, except for Hg and Ag, all observations have been made upon the absorption lines of alkali atoms. The applied pressures of foreign gases have been increased to the order of a thousand atmospheres. For illustration, Fig. 1 is a picture showing the red shift and broadening of the blue doublet of cesium produced by argon, Cs(2)/*A*¶, at 128 atmospheres (262°C). Figure 2 is a plot of the line contours of the same doublet produced by argon and helium respectively.[103]

Customarily, the observations in absorption are carried out with a light source of continuous spectrum, as treated in the previous chapter. However, Kastha,[104] using a mercury lamp as a line source, observed how the Hg 2537 A line shape was modified after it has passed through an absorption cell of mercury vapor at about 100°C. He repeated the same observation with the sodium D line from a sodium lamp through a sodium vapor cell at 200–300°C. It was found that every line appeared separated into two components. Using the vapor pressure-temperature formula, one observes that the separation is linearly proportional to the pressure, and the two components extend to the position at which there was originally no appreciable intensity in the incident light. The change of frequency of the transmitted line may be caused by the absorbing and reemitting mechanism in the absorbing vapor. The intensity diminished with increase in pressure.

Furthermore, the broadening of certain lines caused by pressures of neutral particles can be observed in emission at electrical discharges, provided that those lines, for instance, mercury infrared line[105] (10140 A) or neon (5852 A)[106] are insensitive to the Stark broadening at low current.

1. The Shift of Low Member Lines

A resume of data for the shift** of the *first member* of the principal series (resonance lines) of a number of alkalis produced by *argon* is plotted in Fig. 3. The results for the Na—D lines were taken from Kleman and Lindholm's observations[107] and from Figs. 2 and 3 of Margenau and Watson's paper.[108] Their corrections for "thermal transpiration" are omitted and the shifts are converted into cm^{-1}. Hull's[109] and Ch'en's[110] results are used for the K and Rb resonance lines, respectively. Robin and Robin's[85] results are not drawn on the same scale but are attached as an insert, because (1) their data cover such a wide range of relative density that they could not be included in the figure without seriously reducing the size of the curve as presented, and (2) the precision of their results is rather low, the reciprocal dispersion being 160 A/mm at 7800 A and the errors of measurement being ±5 to ±25 A. For comparison, two of their readings are listed for r.d.†† = 100.

It is to be noted that for small values of r.d., below about 10, the shift increases approximately linearly with r.d. (at a rate of -0.32 cm^{-1}/r.d.). This linear dependence is predicted by the impact theory. For values of r.d. above 19 the curve deviates from linearity and then becomes linear again with about twice the

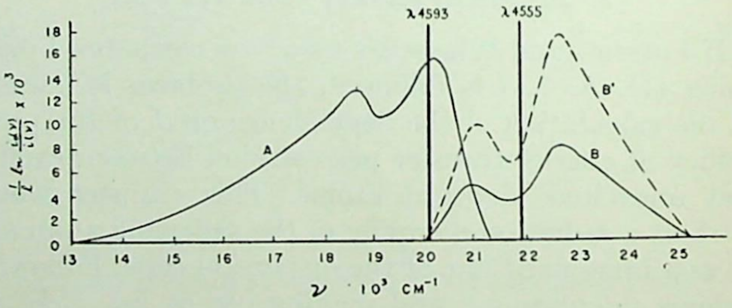


FIG. 2. The broadening and shift of the blue Cs doublet produced by argon and helium. Vertical lines: Position for unperturbed doublet. Curve *A*: for Cs(2)/A, $P=127$ atmos; $T=536$°K. Curve *B*: for Cs(2)/He, $P=54.1$ atmos; $T=519$°K. Curve *B'*: for Cs(2)/He, $P=54.1$ atmos; $T=519$°K, plotted to the same atomic absorption coefficient scale as for Curve *A*. (After Ch'en and Parker[103]).

¶ Cs(2)/A is a notation to indicate the "second" member of the principal series of "cesium" under foreign gas pressure of "argon." This notation will be used very frequently hereafter.

[103] S. Y. Ch'en and W. J. Parker, J. Opt. Soc. Am. 45, 22 (1955).

[104] G. S. Kastha, Indian J. Phys. 23, 24 (1949); 27, 67 (1953).

[105] R. Rompe and Schulz, Z. Physik **108**, 654 (1938); **110**, 223 (1938); P. Schulz, Physik. Z. **38**, 899 (1938).

[106] K. Lang, Acta Phys. Austriaca **5**, 376 (1951); Sitzber. IIa, **161**, 65 (1952).

** The notation for the sign of shift will be adopted as follows: If the shift is given in wavelength, "+" refers to red shift, and "−" refers to violet shift. If the shift is given in wave numbers, "+" refers to violet, and "−" refers to red shift. However, when a comparison of the amount of shift is described, its "absolute value" is always implied.

[107] B. Kleman and E. Lindholm, Arkiv Mat. Astron. Phys. **32B**, No. 4, Paper 10 (1946).

[108] H. Margenau and W. W. Watson, Phys. Rev. **44**, 92 (1933).

[109] G. F. Hull, Jr., Phys. Rev. **50**, 1148 (1936).

[110] S. Y. Ch'en, Phys. Rev. **58**, 1051 (1940).

†† Relative density will be hereafter abbreviated as r.d., the unit being the density of the same quantity of a foreign gas as that used in the experiment but at 0°C and 1 atmosphere.

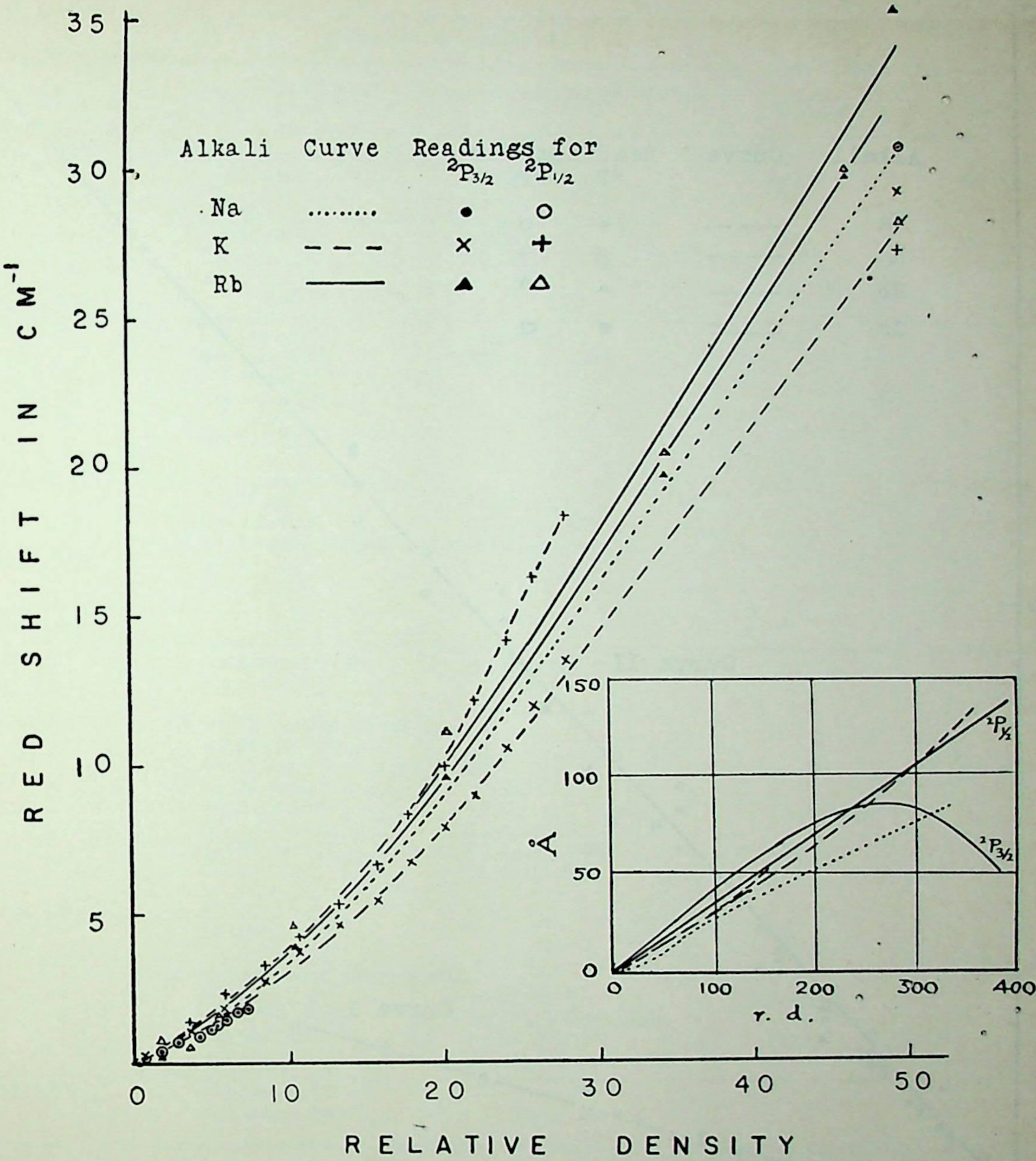


FIG. 3. The shift of the resonance lines of alkali metals produced by argon. Note: The ordinate for the main curve is shift in cm^{-1}, while that for the insert is shift in A. The abscissas are the relative densities of argon.

former slope (-0.71 cm^{-1}/r.d.) for r.d. above 25. Since it is almost impossible at present to get a correct idea of the atomic interaction forces for the excited states involved, a test of the theories as described in Part II can hardly be made. But, according to Margenau's rough approximation, the transition point is assumed to take place where the linear collision theory, Eq. (57), would give the same value for $\Delta\nu_m$ as is given by the quadratic statistical theory, Eq. (97), under the dispersion interaction alone, namely

$$N=\frac{12.88}{0.822\pi^3}\left(\frac{kT}{2m}\right)^{3/10}\cdot|\gamma|^{-3/5}. \qquad (134)$$

γ can be evaluated through Eq. (3). The transition point N should be somewhere near 20 atmospheres for these cases as observed, and should move toward higher r.d. for smaller γ.

However, the shift does not show a quadratic relationship with r.d. when the pressure is high. According to the results of Robin and Robin the shift is essentially linear up to about r.d. 400 for all observed lines except the $^2P_{\frac{3}{2}}$ component of Rb. In the latter case the shift increases linearly for values of r.d. below 150, becomes a maximum at r.d. about 260, and then decreases for values of r.d. above this figure.

The shape of the shift *vs* r.d. curve for Hg λ2537

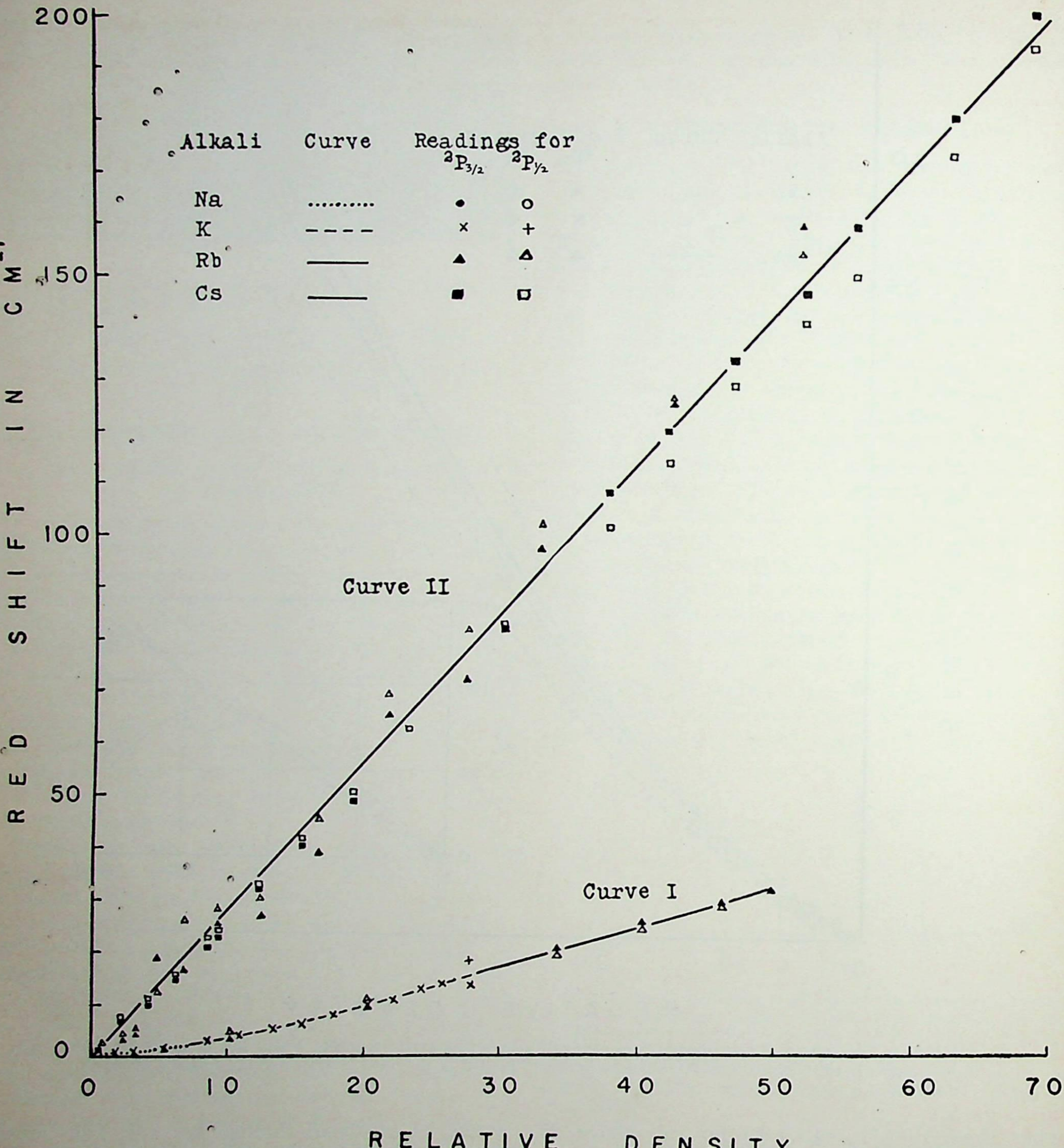


FIG. 4. Curve II is a plot of the shift of the second doublet of the absorption series of alkali metals (Rb and Cs) produced by argon; Curve I the corresponding curve for the first doublet plotted for comparison.

produced by argon is essentially the same as that for the resonance lines of the alkalis, except that the shift is about half as large in cm^{-1}, or twenty times smaller in A. The shift is -0.127 cm^{-1}/r.d.[111] for r.d.<50, and -0.323 cm^{-1}/r.d.[92] for $200<$r.d.<300. This relatively smaller shift for Hg λ2537 may be attributed to the small f value of the line in comparison with the f value of the alkali resonance lines.

The shift of Ag 3280, 3382 is -0.284 cm^{-1}/r.d. for r.d.<18.9,[112] and that of Pb 2833 is -0.5 cm^{-1}/r.d. for r.d.<8.[110]

A similar set of results for the shift of the second doublet of the absorption series of alkalis produced by

[111] Füchtbauer, Joos, and Dinkelacker, Ann. Physik 71, 204 (1923).

[112] E. D. Clayton and S. Y. Ch'en, Phys. Rev. 85, 68 (1952).

TABLE I. The mean shift of various low member lines per unit relative density, r.d., of various foreign gases under different ranges of r.d. values.

Absorption line/foreign gas	Mean shift/r.d. for low r.d. range $\langle\Delta\nu'\rangle$	Mean shift/r.d. for high r.d. range $\langle\Delta\nu''\rangle$	Ratio $\langle\Delta\nu''/\Delta\nu'\rangle$	Reference
(1) Na, K, Rb/A	(i) for resonance lines, $\langle\Delta\nu\rangle_1$			a
	-0.32 cm^{-1} (for r.d. <10)	-0.71 cm^{-1} (25 <r.d. <200)	2.22	See Fig. 3
	(ii) for the second member of absorption series, $\langle\Delta\nu_2\rangle$			
	-2.81 cm^{-1} (for r.d. <70)	-6.03 cm^{-2} (120 <r.d. <300)	2.14	Fig. 4
Ratio $\langle\Delta\nu\rangle_2/\langle\Delta\nu\rangle_1$	8.80	8.50		
(2) Hg 2537/A	-0.149 cm^{-1} (for r.d. <50)	-0.323 cm^{-1} (200 <r.d. <300)	2.54	b
Ag 3280/A, 3382/A	-0.28 cm^{-1} (for r.d. <19)			c
Pb 2833/A	-0.5 cm^{-1} (for r.d. <8)			
(3) Na, K, Rb/N$_2$	(i) for resonance lines, $\langle\Delta\nu_1\rangle$		2.5	See Fig. 5
	-0.27 cm^{-1} (for r.d. <15)	-0.67 cm^{-1} (20 <r.d. <50)		
	(ii) for the second member of absorption series, $\langle\Delta\nu_2\rangle$			
K, Rb, Cs/N$_2$	-0.53 cm^{-1} (for r.d. <8)			
Ratio $\langle\Delta\nu\rangle_2/\langle\Delta\nu\rangle_1$	2.0			
Hg/N$_2$	-0.124 cm^{-1}/r.d. (for r.d. <50)	-0.139 cm^{-1} (r.d. 100)		b
(4) Rb, Cs/He	(i) for Rb resonance lines, $\langle\Delta\nu\rangle_1$			d
	$+0.24$ cm^{-1} ($^2P_{\frac{1}{2}}$)			
	$+0.13$ cm^{-1} ($^2P_{\frac{3}{2}}$) (for r.d. <50)			See Fig. 6
	(ii) for the second member of Rb, $\langle\Delta\nu_2\rangle$			e
	$+1.3$ cm^{-1} ($^2P_{\frac{1}{2}}$) (for r.d. <12)	$+3.6$(?) cm^{-1} ($^2P_{\frac{1}{2}}$) (25 <r.d. <50)	2.8(?) ($^2P_{\frac{1}{2}}$)	
	$+0.53$ cm^{-1} ($^2P_{\frac{3}{2}}$) (for r.d. <12)	$+2.4$ cm^{-1} ($^2P_{\frac{3}{2}}$) (25 <r.d. <50)	4.5 ($^2P_{\frac{3}{2}}$)	
Ratio $\langle\Delta\nu\rangle_2/\langle\Delta\nu\rangle_1$	5.4 ($^2P_{\frac{1}{2}}$) 4.1? ($^2P_{\frac{3}{2}}$)			
	(iii) for the second member of Cs			f
	$+1.4$ cm^{-1} (?) (for $^2P_{\frac{1}{2}}$, r.d. <8)	$+3.51$ cm^{-1} (for $^2P_{\frac{1}{2}}$, 9 <r.d. <40)	2.5(?)	
	$+0.583$ cm^{-1} (for $^2P_{\frac{3}{2}}$, r.d. <12)	$+4.65$ (?) cm^{-1} (for $^2P_{\frac{3}{2}}$, 18 <r.d. <40)	7.9(?)	
(5) Hg 2537/He	$+0.079$ cm^{-1} (r.d. = 0 to 500)			g

[a] See references 82, 85, 107–110.
[b] See references 84, 111.
[c] See reference 112.
[d] See reference 110.
[e] See reference 113.
[f] See reference 103.
[g] See reference 86.

argon is shown in Fig. 4. The results are taken from those of Ch'en and his associates.[103,113] The slope of the curve is -2.81 cm^{-1}/r.d. Robin and Robin[85] observed the shift of the second doublet of the principal series of Rb perturbed by argon for r.d. up to 300 with somewhat lower precision (± 2 to ± 10 A). The two components of the doublet fused into a single band as a result of broadening at pressures higher than r.d. 70. Their curve for r.d. less than 70 agrees very well with Ch'en's result. At r.d. about 90 the slope of the curve is practically doubled (-6.03 cm^{-1}/r.d. for 120 < r.d. < 300).

It is apparent at once that for the first two doublets of the alkali absorption series, the shift (in cm^{-1}) is chiefly dependent on the member of the series and on the nature of the foreign gas but is essentially the same for all alkalis (with exceptions at high pressures, r.d. > 200, at which the anomaly for the $^2P_{\frac{1}{2}}$ component of the Rb resonance ines is observed). As shown in Table I, the slope of the curve for the second doublet is about 8.6 times greater than that for the first doublet. The curves for argon are very straight up to r.d. 10 and 70 for the first and second doublets respectively (and up to r.d. 50 for Hg 2537). Above these respective r.d. the curves reach slopes something more than double the value for low r.d.

The experimental observations as described in the previous paragraph indicate that the force laws for the

[113] S. Y. Ch'en and D. A. Kohler, Phys. Rev. 90, 1019 (1953).

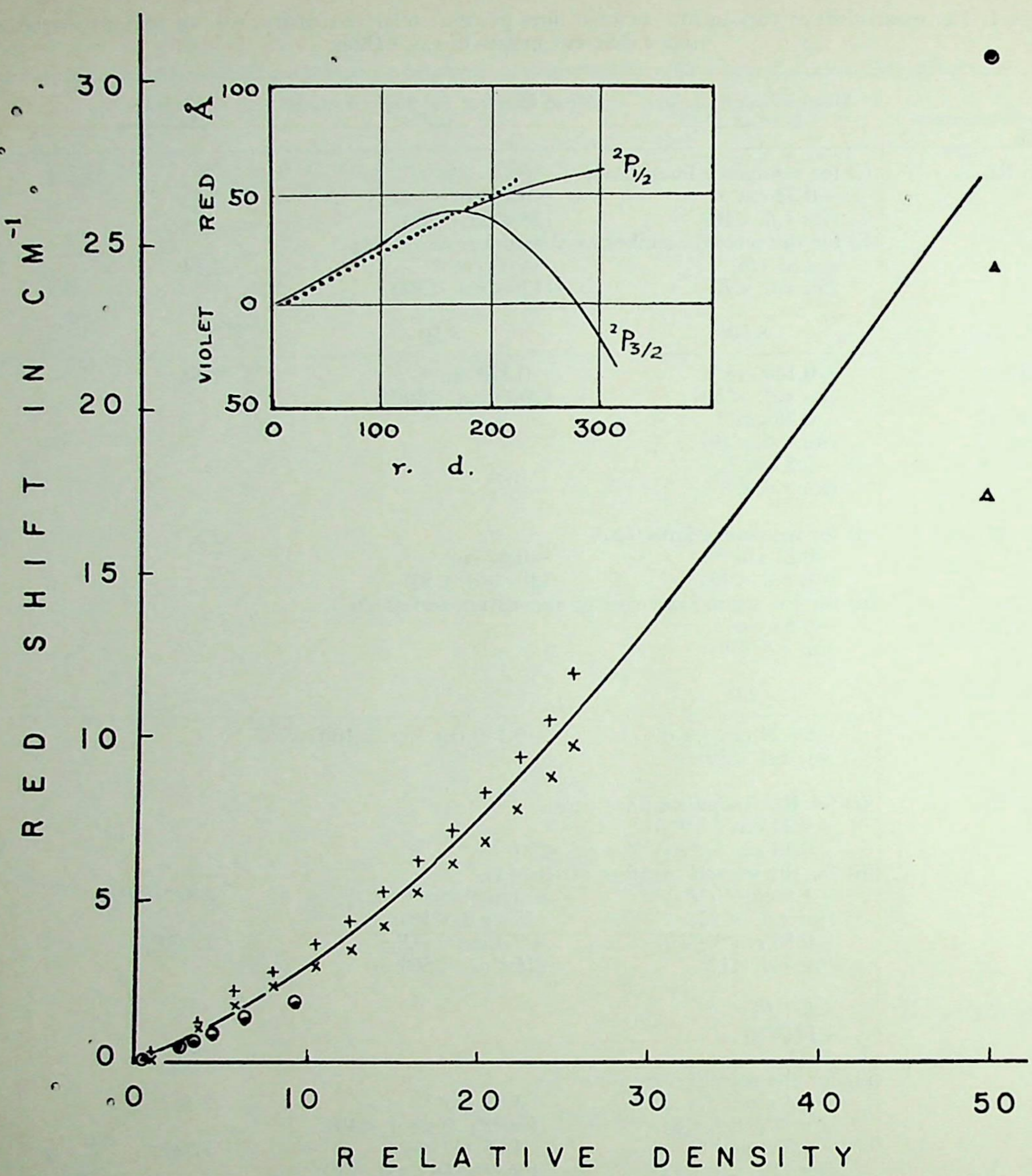


FIG. 5. The shift of the resonance lines of alkali atoms produced by nitrogen, plotted in the same way as in Fig. 3.

excited states of the two series lines are different. Because of the fact that the shift of the second member is greater than that of the first member, the γ value for the excited state of the second member should be greater. According to Eq. (134) the transition point for the second member would be smaller if the same type of dispersion forces are assumed for both members.

Quite plentiful data are also available for *nitrogen* and helium. As shown in Fig. 5, N_2 also produces a red shift and the curve swings up slightly at r.d. around 20. The shift for N_2 is a little less than that for argon. The points for the Na−D lines are taken from Margenau and Watson[108] and those for the K resonance lines from Hull.[109] The points in the figure for r.d. 50 were obtained by interpolation from Robin's result,[84] which is again shown as an insert. It is not easy to observe the pressure effect of N_2 or H_2 when the pressure is higher than 30 atmospheres because the alkali will start to react with the gas.[110] Data for the second member of the principal series can be obtained from Watson and Margenau's[114] result for K, from Ch'en and Pao's[115] result for Rb, and from Füchtbauer and Gössler's[116] result for Cs. The shift of the second member [−0.60 and −0.45 cm^{-1}/r.d. for K(2)/N_2 and Rb(2)/N_2, respectively] is twice that for the first member of the principal series. All authors reported that the shift of the third member is less than that of the second member by about 40% for Na, 30% for K, and 60% for Cs. According to Robin's result, as shown in the inserts of Fig. 5, as the pressure of N_2 increases the shorter wavelength ($^2P_{\frac{3}{2}}$) component of the Rb reso-

[114] W. W. Watson and H. Margenau, Phys. Rev. 44, 748 (1933).
[115] S. Y. Ch'en and C. S. Pao, Phys. Rev. 58, 1058 (1940).
[116] C. Füchtbauer and F. Gössler, Z. Physik 87, 89 (1934).

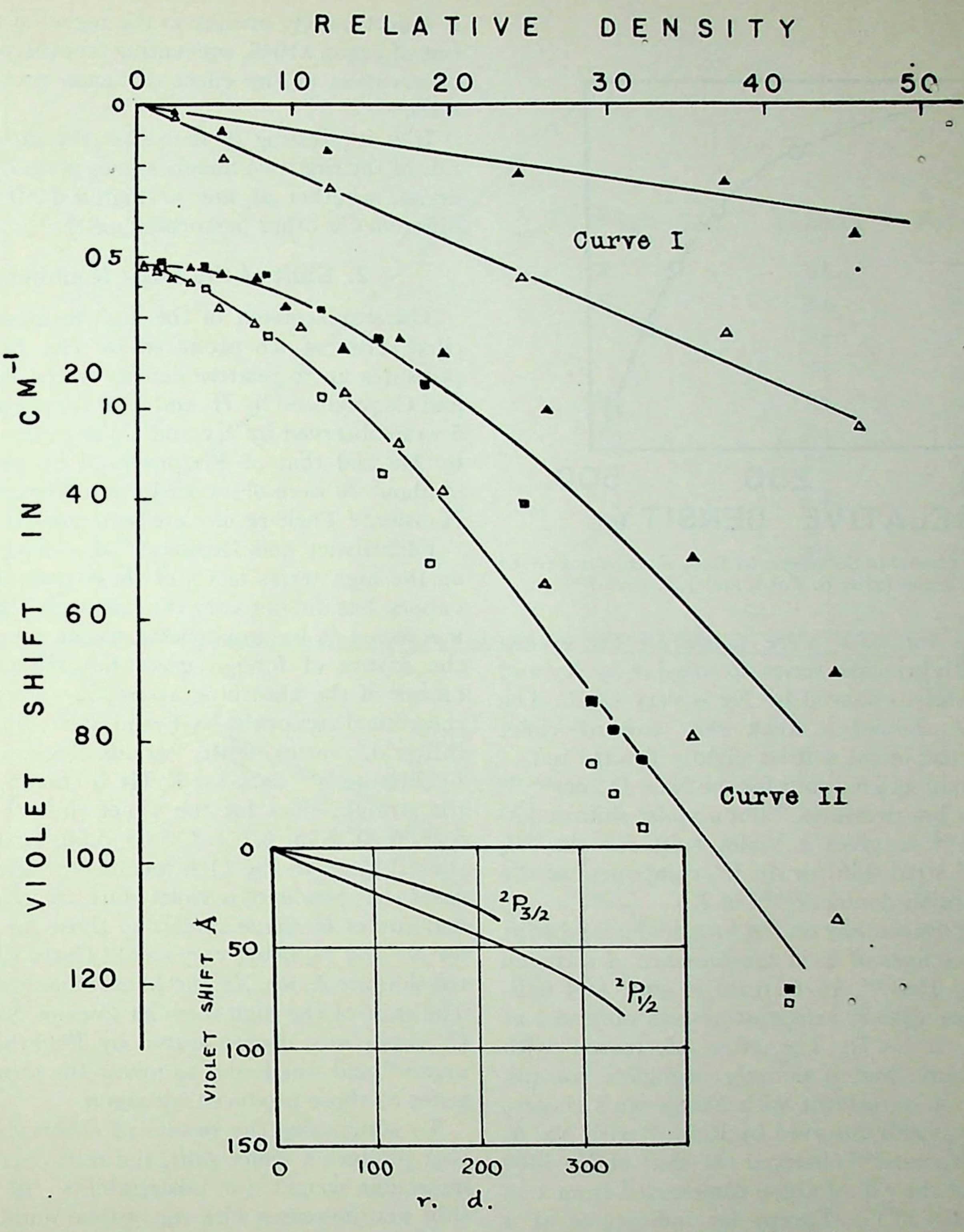


FIG. 6. The shift of the first doublet (Curve I) and of the second doublet (Curve II) of the absorption series of alkali atoms (Rb and Cs) produced by helium. The scale of the ordinate for Curve I is five times that for Curve II. The insert is the shift in A *vs* r.d. plot of the Rb resonance lines produced by He.

nance line first shifts towards *red*, then, at r.d. around 160, the shift swings back quite rapidly, and then, when the r.d. was over 270, the shift turns to *violet* with respect to the position of the line with no foreign gas pressure. This would lead us to believe that when the pressure is sufficiently high, say many thousand atmospheres, the shift will be toward violet for all gases, at least for all resonance lines. (For higher member lines, the potential curves of the excited states could cross each other at some magnitude of interactions. The distortion of levels due to crossing will prevent us from foreseeing, without detailed calculation, whether the shift will always be toward the violet at very high pressures.)

Helium produces a strong violet shift, as shown in Fig. 6 and Table I. The results are those observed by Ch'en and his associates[103,110,113] for Rb and Cs. The shift of the second doublet is about five times greater than that for the first doublet. Füchtbauer and Gössler observed that there was a switch from red to violet shift for Cs(1)/He to Cs(2)/He, and for Cs(2)/Ne to Cs(3)/Ne. The shift of Hg 2537 by helium was also toward violet with a value of $+0.031$ cm^{-1}/r.d.[86b] to $+0.079$ cm^{-1}/r.d.[86a] (-0.0051 A/r.d.).

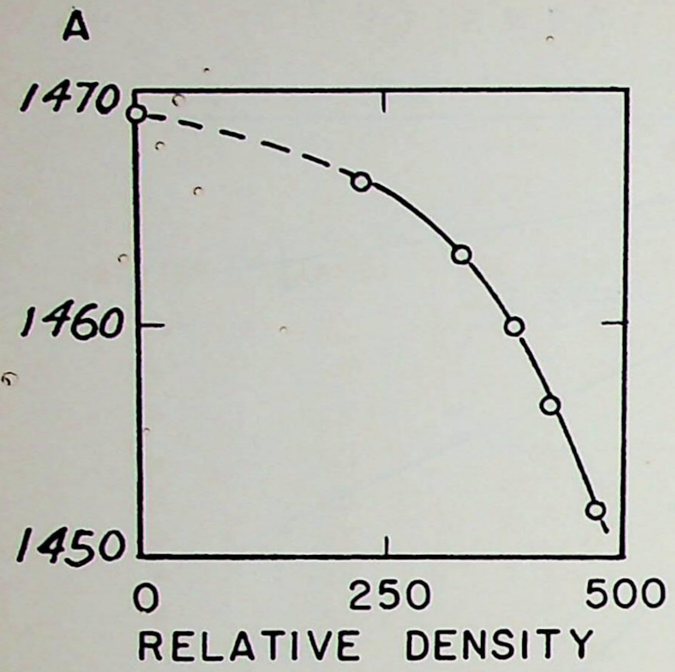


FIG. 7. The displacement of the resonance lines of xenon due to the presence of argon (after S. Robin and J. Romand[120]).

Observations for *neon* were made on the *second doublet* of the Rb principal series up to r.d. 8 by Ny and Ch'en.[117] The shift produced by Ne is very small. The $^2P_{\frac{1}{2}}$ component showed a weak shift toward violet while the $^2P_{\frac{3}{2}}$ component shifted slightly toward red.

Hydrogen produces a red shift for the Na−D lines[107,108] and Hg 2537 at low pressures,[85] but a violet shift at 120 atmospheres.[86,118] It gives a violet shift for the $^2P_{\frac{1}{2}}$ component and a red shift for the $^2P_{\frac{3}{2}}$ component of the first two absorption doublets[110,115] of Rb.

The effect of *temperature* on the broadening and shift of the resonance lines of K in the presence of nitrogen was studied by Hull.[119] An increase in shift and half-width due to increase in temperature was observed at low pressures (r.d. <7). The effect decreased with increased pressure, and is entirely negligible for r.d. higher than 9, in agreement with Margenau's theory. The last fact was also observed by Robin[83] with Na/A.

Robin and Romand[120] observed the shift of Xe 1469 as a function of the r.d. of argon compressed from 1 to 800 kg/cm² ($T=20°$C). Except for indications of a slight red shift at low pressures, the shift was about −3 A (toward the violet) for r.d. 200, and then increased rapidly as further pressure was added. The results are shown in Fig. 7. It is particularly interesting that argon produced *violet* shifts and *violet* asymmetry for Xe λ1469 in contrast to all other observations with argon. The different behavior in this case, in contrast to the effects observed with alkali atoms, may be due in part to the fact that both argon and xenon have completed electron shells in the ground states. Xenon is unfortunately opaque in the region of the absorption line of argon λ1048, preventing an otherwise interesting observation of the effect of xenon pressure on argon lines.

It is interesting to note that the ratio of shifts per r.d. of the first two member lines is nearly the same for argon, whether at low or high r.d., but the ratio is different for other perturbing gases.

2. Shift of the High Member Lines

The displacement of the high members of the principal series of Rb produced by He, Ne, and A[121] for pressures up to relative density 6 and that of Na, Rb, and Cs produced by H_2 and N_2[122] for pressures up to r.d. 5 were observed by Ny and Ch'en; that of K produced by He and that of Rb produced by neon, up to r.d. of about 20 were observed by Füchtbauer, Heesen, and Hansler.[123] Their results are summarized in Fig. 8.

Füchtbauer and Reimers[124] also observed the effect on the high terms of Cs of these gases and of paraffin vapors, but did not vary the pressure. The displacement was found to be proportional to the r.d., dependent on the nature of foreign gases but independent of the nature of the absorbing atoms, in agreement with the theoretical treatment by Fermi.[125] To illustrate how the shift/r.d. varies with various high member lines, Füchtbauer's[123] data for K/He is chosen. The slopes of the straight lines for the violet shift *vs* r.d. plots are 5.25, 6.20, 5.75, 6.2, 6.2, 5.63, 5.60, 5.50 cm⁻¹/r.d. for the 4th‡‡ up to the 11th members, respectively. Other gases that produced a violet shift, *viz.*, H_2, Ne, and N_2, had curves of shape similar to those for He, the shifts for Ne and N_2 being very small. Gases that produced a red shift are A, Kr, Xe, methane, ethane, and propane.[124] The shift of the high member lines of Na displaced by Cs vapor was alsome asured by Füchtbauer and Heimann[126] and was found to follow the same trend in the series as those produced by argon.

To summarize the results of observations, for gases that produce a violet shift, the shift decreases with the molecular weight (or polarizability) of the gas. The shift first increases with the ordinal number of the lines in the series, then attains a weak maximum at the 5th to 8th members, and finally approaches a constant value for lines higher than the 15th member up to the limit of the series.[121,122] For gases that produce a *red* shift, the shift *increases* with the molecular weight (or polarizability) of the gas. The shift *vs* series member plots have the same shape as those in the first case but there is *no* such weak maximum around the 5th–8th

117 T. Z. Ny and S. Y. Ch'en, Phys. Rev. **52**, 1158 (1937).
118 This change in direction of shift was also observed by J. Granier-Mayence and S. Robin, J. phys. radium **13**, 494 (1952), with the No band perturbed by A.
119 G. F. Hull, Jr., Phys. Rev. **51**, 572 (1937).
120 S. Robin and J. Romand, Compt. rend. **231**, 1455 (1950).

121 T. Z. Ny and S. Y. Ch'en, Phys. Rev. **51**, 567 (1937).
122 T. Z. Ny and S. Y. Ch'en, Phys. Rev. **54**, 1045 (1938).
123 C. Füchtbauer and W. V. Heesen, Z. Physik **113**, (1939); and reference 130.
124 C. Füchtbauer and H. J. Reimers, Z. Physik **97**, 1 (1935).
125 E. Fermi, Nuovo cimento **11**, 157 (1934).
‡‡ There might be something erroneous about their[82] data on the shift of the $1s-5p$ line, as given on page 327 of that paper, if the line was not shifted towards red for r.d. less than 2.
126 C. Füchtbauer and G. Heimann, Z. Physik **110**, 8 (1938).

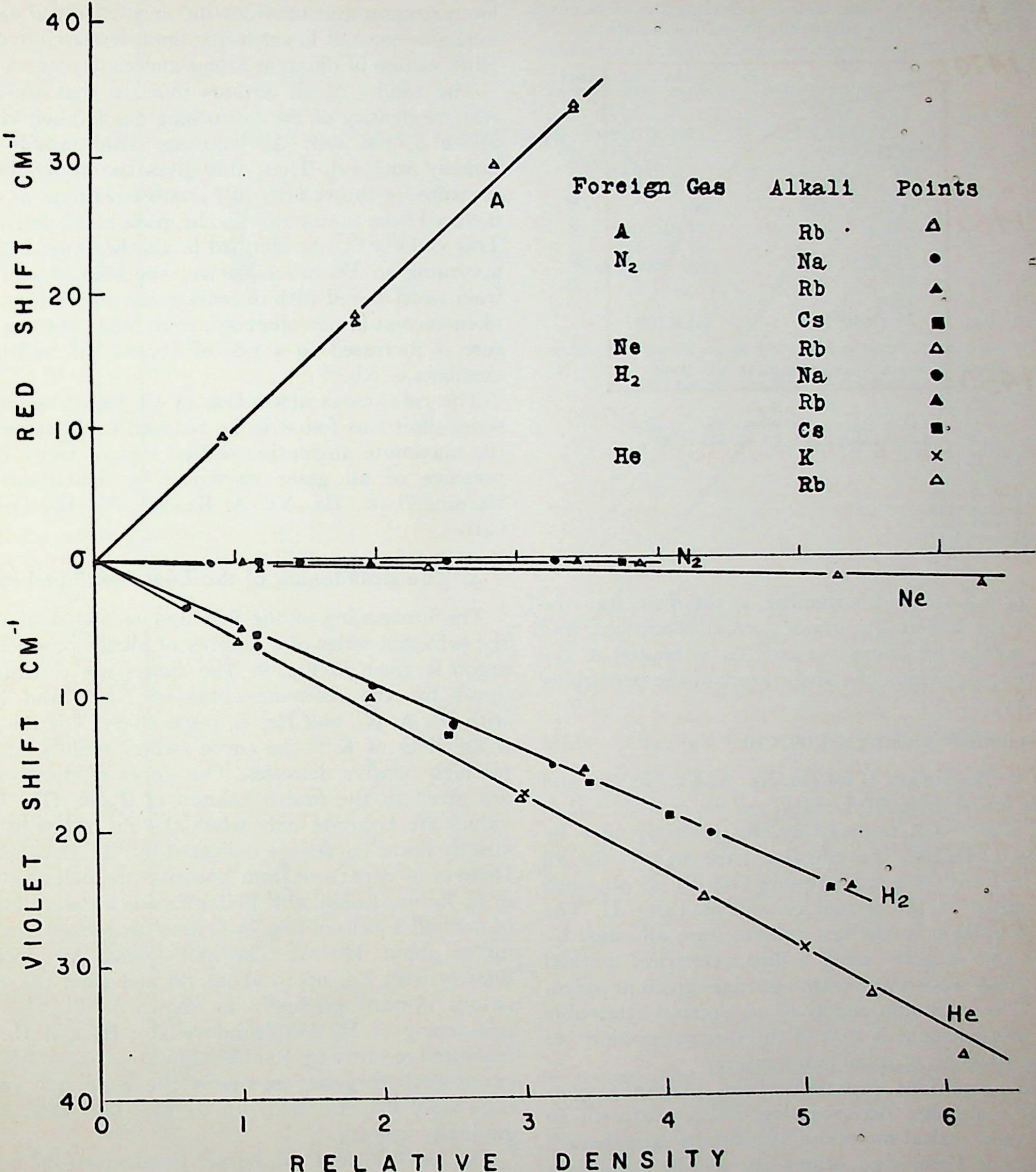


FIG. 8. The mean shift of the high members of the principal series of alkali atoms produced by various gases.

members. According to Fig. 8, the observed shift (in cm^{-1}/r.d.) for A, N_2, Ne, H_2, and He are -9.7, $+0.1$, $+0.2$, $+4.7$, and $+5.9$, respectively.

According to Fermi's theory§§ the effective collision cross section of single atoms of perturbing gases for electrons at very low relative velocity can be deduced from observation of the constant shifts of the high series lines of the alkalis at low pressure. The method is to take the difference between the observed shift Δ and the shift, Δ_ϵ, due to the polarization of the foreign atoms in the field of the core of the alkali atom to obtain Δ_σ, the shift caused by the perturbation of the valence electron by the foreign atoms.

$$\Delta_\epsilon = 20e^2(\epsilon_0 - 1)N^{4/3}/8\pi hcL_0$$
$$= 0.000922(\epsilon_0 - 1)N^{4/3}/L_0 \text{ cm}^{-1}, \quad (135)$$

§§ Sampson and Margenau have adapted this theory to the broadening of impurity levels in semiconductors, which arises from collisions between a hole and the phonons. D. Sampson and H. Margenau, Phys. Rev. **103**, 879 (1956)

TABLE II. The effective cross section of certain gases for electrons of very low velocity as measured by different authors.

Gas	By optical method, cm^2/cm^3	By electrical method, cm^2/cm^3 (kinetic energy of electron)
He	11.7[a], 15.5,[b] 15.5[c]	15.3 (0.03 ev)[i]
Ne	0.23,[b], 0.24[c]	3.13 (0.2 ev)[j]
A	42,[a] 25.2,[b] 23.7[c]	2.6 (0.03 ev)[i]
H_2	12.3,[a] 14.0[d]	24.4[k]
N_2	0.93,[a] 5,[b] 2.0[d]	10[k]
Kr	121.3[e]	20 (0.6 ev)[l]
Xe	412[f]	58 (0.2 ev)[l]
Hg	35–37.2[f]	250–300 (1 ev)[m]
CH_4	73[g]	
C_2H_6	120[g]	
Cs on Na	25.9×10^{3}[h]	1.1×10^{3}[n]

[a] E. Amaldi and E. Segré, Nuovo cimento 11, 145 (1934).
[b] Füchtbauer, Schultz, and Brandt, Z. Physik 90, 403 (1934).
[c] See reference 121.
[d] See reference 122.
[e] C. Füchtbauer and H. J. Reimers, Z. Physik 95, 1 (1935).
[f] C. Füchtbauer and F. Gössler, Z. Physik 93, 648 (1935).
[g] See reference 124.
[h] See reference 126.
[i] See reference 127a.
[j] See reference 127b.
[k] See reference 114a.
[l] See reference 127c.
[m] See reference 127d.
[n] See reference 127e.

where L_0 is Loschmidt's number, ϵ_0 the dielectric constant of the foreign gas under normal conditions, and N the number of atoms per cm^3. Δ_σ is related to the effective cross section of a single atom of the perturbing gas, σ, by

$$\Delta_\sigma = \pm hN\sigma^{\frac{1}{2}}/4\pi^{\frac{3}{2}}mc = \pm 1.09\times10^{-11}N\sigma^{\frac{1}{2}}\ \text{cm}^{-1}. \quad (136)$$

The Δ_σ values (in cm^{-1}) for He, Ne, A, Kr, Xe, H_2, and N_2 were found to be +6.1, +0.7, −7.95, −17.2, −31.6, +5.85, and +2.2, respectively. By multiplying σ by $L_0/760 = 3.54\times10^{16}$, the effective cross section for all atoms in 1 cm^3 at 1 mm pressure and 0°C is obtained as tabulated in the second column of Table II. The values tabulated in the last column were obtained by Wahlin and others[127] using a direct electrical method for the cross section at electron energies given in parentheses. The minimum energy of an electron attainable in the measurement is that of the thermal motion per electron (0.03 ev at room temperature).

Fermi's equations were derived for spherically symmetrical molecules. An extension of his treatment to axially symmetrical molecules as given by Reinsberg[128] showed that Fermi's equations can also be applied to them without any appreciable correction. Furthermore, Firsov[129] repeated the calculation in a more general way and found that Fermi's theory is good for shift. In view of the strong dependence of the cross section on electron velocity, Fermi's theory is satisfactory at low pressures and provides the only values at present available for the low-velocity limit of elastic collision cross section of different atoms and molecules.

The results of all authors indicate that when the relative density of the perturbing gas is high, namely above 5, the shift Δ_σ increases more rapidly than linearly with r.d. Thus, this gives the upper limit in pressure for the validity of Fermi's equations, as easily deduced from the assumption he made in the derivation. This validity can be clarified in another way. Namely, according to Fermi's equation, the shift should pass from violet to red with increasing relative density. This phenomenon is not observed, even when the Ne pressure is increased to a r.d. of 16, on the high series members of Rb.[130]

The continuous absorption of Cs vapor beyond its series limit was found to be reduced in intensity, and the maximum absorption shifted toward violet in the presence of all gases employed by Ditchburn and Harding,[131] *viz.* He, Ne, A, Kr, Xe, N_2, H_2, D_2, and C_6H_6.

3. The Broadening of the Low Member Lines

The broadening of the first and second doublets of the principal series of a number of alkalis produced by argon is given in Fig. 9. The curves are consistently linear for the resonance lines of Na[108] and Rb[110] with H_2, A, N_2, and He, A, respectively. For the resonance lines of K[109] the curve swings slightly upward for high relative densities. The slopes of these curves are given in the fourth column of Table III. These values are accurate only when the r.d. is low.‖‖ The strictly linear curves are indicated by "L" in column 4. Degrees of departure from linearity are indicated by l or l-. Robin, Robin, and Vodar[82] made a determination of the half-widths of the Na-D lines broadened by argon up to about 340 r.d. The half-widths, $\Delta\nu_{\frac{1}{2}}$, increase linearly with r.d. up to about 80 and then the curve swings upward gradually as shown in Fig. 10. The broadening of Hg 2537 produced by H_2 and He was measured recently by Robin.[86] The $\Delta\nu_{\frac{1}{2}}$ (in cm^{-1}) *vs* r.d. curve for both gases are nearly the same, and have a slope of 0.363 cm^{-1}/r.d. up to r.d. 430, then swing gradually upward.

The half-width *vs* r.d. curves obtained so far for the *second member* of the principal series of all alkalis are linear for low r.d., say up to 8, and for all gases. When the r.d. is high, the curve for argon still follows the linear relationship as before, while for helium the curve indicates a definite upward swing.[103,113] In plotting Curve II of Fig. 11, only readings for the $^2P_{\frac{1}{2}}$ component were used. The curve for the $^2P_{\frac{3}{2}}$ component was left

[127] (a) H. B. Wahlin, Phys. Rev. 37, 260 (1931); (b) C. Normand, *ibid.* 35, 1217 (1930); (c) C. Ramsauer and R. Kollath, Ann. Physik 3, 536 (1929); 4, 91 (1930); (d) R. B. Brode, Proc. Roy. Soc. (London) A109, 397 (1925); (e) R. B. Brode, Phys. Rev. 33, 1069 (1929).
[128] (a) C. Reinsberg, Z. Physik 93, 416 (1935); (b) C. Reinsberg, Z. Physik 105, 460 (1937).
[129] O. B. Firsov, Doklady, Akad Nauk S.S.S.R. 61, 357 (1949); Zhur. Eksptl. Theort. Fiz. 21, 627 and 634 (1951).

[130] C. Füchtbauer and G. Hansler, Physik. Z. 41, 555 (1940).
[131] R. W. Ditchburn and J. Harding, Proc. Roy. Soc. (London) A157, 66 (1936).
‖‖ When the r.d. was high, say above 10, the slopes were usually higher (see reference 116 and 117 in comparison with references 103 and 113).

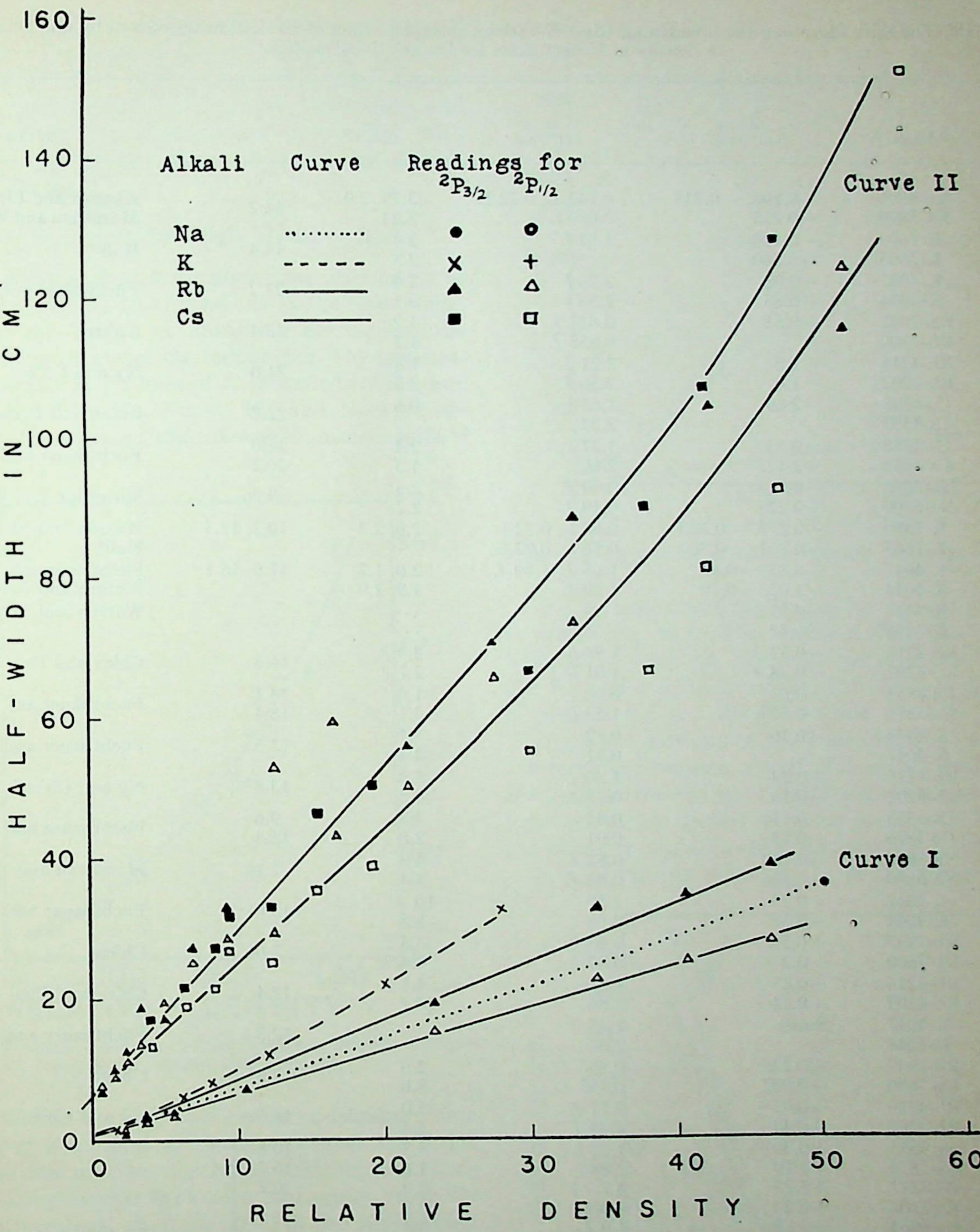


FIG. 9. Pressure broadening of the first (Curve I) and the second (Curve II) doublets of the principal series of alkali atoms produced by argon.

out because of its overlapping with the Rb—He diffuse band appearing at the short wavelength side.[103] The $\Delta\nu_{\frac{1}{2}}$ *vs* r.d. plot for Hg (2537)/He is[86b] linear up to r.d. 110 approximately.

The violet asymmetrical broadening of the resonance lines of K under the vapor pressure of Rb (150 mm Hg) was observed by Ch'en.[132] There is no measurable shift.

The temperature effect on line widths is much complicated. Theoretically, it is given by Eq. (37) (Lorentz) as $\sqrt{T}$ or by (57) as $T^{3/10}$ with N kept constant at low pressures. However, experimentally, the confirmation of either of them has not been established. Orthman[133] observed $\sqrt{T}$ effect with Hg 2537/H_2, while Horodoniczy and Jablonski,[134] very small T effect with Hg 2537/He

[132] S. Y. Ch'en, Phys. Rev. **73**, 1470 (1948).

[133] W. Orthman, Ann. Physik **78**, 601 (1925).

[134] H. Horodnicsy and A. Jablonski, Nature **142**, 1122 (1938); **144**, 594 (1939).

TABLE III. The shift ($\Delta\nu_m$) and the broadening ($\Delta\nu_{\frac{1}{2}}$) of various absorption lines of alkali atoms produced by unit relative density of foreign gases for low relative densities.

Gas	Lines	$\Delta\nu_m$ cm^{-1}/r.d.	$\Delta\nu_{1/2}$ cm^{-1}/r.d.	$\Delta\nu_{1/2}/\Delta\nu_m$	ρ Å	Authors
A	Na-5896	−0.196, −0.248	0.742 L, 0.72 L	3.79, 2.9	11.2	Kleman and Lindholm[a]
	Na-5890	−0.213	0.689 L	3.23		Margenau and Watson[b]
A	K-7699	−0.420	1.01 *l*	2.4	14.4	Hull[c]
	K-7665	−0.360		2.8		
A	K-4047	−0.92	2.20 *l*	2.4	22.2	Füchtbauer and Reimers[d]
	K-4044	−0.83	2.58 *l*	3.1		
A	Rb-7947	−0.51 ?	0.627 *L*	1.2	13.4	Ch'en[e]
	Rb-7800		0.855 *L*	1.7		
A	Rb-4216	−1.2	2.21 *L*	1.8	24.0	Ny and Ch'en[f]
	Rb-4202	−1.0	2.56 *L*	2.6		
A	Cs-4593	−2.69	1.63 *l*	0.6	22.2	Ch'en and Parker[g]
	Cs-4555		2.22 *l*			
A	Cs-4555	−0.53	1.27 *l*	2.4	18.0	Füchtbauer and Gössler[h]
	Cs-3876	−2.12	2.69	1.3	26.2	
N_2	Na-5896	−0.214	0.49 *L*	2.3	8.96	Margenau and Watson[b]
	Na-5890	−0.223	0.49 *L*	2.2		
N_2	K-7699	−0.277, −0.360	0.56 *l*-, 0.82 *l*-	2.0, 2.3	10.2, 12.3	Watson and Margenau,[i]
	K-7665	−0.261, −0.300	0.56 *l*-, 0.82 *l*-			Hull[c]
N_2	K-4047	−0.83, −0.81	1.65 *L*, 1.39 *L*	2.0, 1.7	17.5, 16.1	Füchtbauer and Reimers[d]
	K-4044	−1.02, −0.70	1.89 *L*	1.9, 2.0		Watson and Margenau[i]
N_2	K-3447	−0.57	...	...		Watson and Margenau[i]
	K-3446	−0.62	...	...		
N_2	Rb-4216	−0.52 ?	1.51 *l*-	2.9 ?	16.3	Ch'en and Pao[j]
	Rb-4202	−0.38 ?	1.01 *l*-	2.7 ?		
N_2	Cs-4555	−0.58	0.90 *l*	1.6	14.1	Füchtbauer and Gössler[h]
	Cs-3876	−0.50	1.53 *l*	3.1	18.4	
Ne	K-4044	−0.16	0.82	5.1	11.3	Füchtbauer and Reimers[d]
	K-4047		0.71	4.4		
Ne	Rb-4216	+0.22	1.30 *L*	5.9	13.8	Ny and Ch'en[f]
	Rb-4202	−0.16	0.73 *L*	4.5		
Ne	Cs-4555	−0.11	0.47	4.3	9.6	Füchtbauer and Gössler[h]
	Cs-3876	−0.33	0.91	2.8	13.3	
H_2	Na-5896	−0.189	0.82 *L*	4.3	7.16	Margenau and Watson[b]
	Na-5890	−0.186	0.82 *L*	4.4		
H_2	K-4047	−0.23	2.39	10	13.0	Füchtbauer and Reimers[d]
	K-4044	−0.33	2.88	8.5		
H_2	Rb-7947	+0.2 ?	0.65 ?	3.5 ?	7.01 ?	Ch'en[e]
	Rb-7800	−0.2 ?	0.87 ?	4.4 ?		
H_2	Rb-4216	+0.45 ?	1.87	4.1 ?	12.4	Ch'en and Pao[j]
	Rb-4202	−0.14 ?	2.86	20 ?		
He	K-4047	Small	1.45		12.3	Füchtbauer and Reimers[d]
	K-4044		2.02			
He	Rb-7947	+0.228	0.595 *L*	2.6	7.8	Ch'en[e]
	Rb-7800	+0.092	0.735 *L*	8.0		
He	Rb-4216	+0.93	2.77 *L*	3.0	14.5	Ny and Ch'en[f]
	Rb-4202	+0.43	1.88 *L*	3.4		
He	Cs-4555	+0.39	1.28 *l*	3.3	10.8	Füchtbauer and Heesen[k]
He	Cs-3876	+1.79	2.88 *l*	1.6	16.2	
He	Cs-3217	+5.25	4.7 ? *l*-	0.9	20.7 ?	
He	Cs-3102	+6.20	2.8 ? *l*	0.5		
He	Cs-3034	+5.75	1.6 *L*-	0.3		
He	Cs-2992	+6.2	1.6 *L*	0.3		
He	Cs-2963	+6.2	1.5 *L*	0.2		
He	Cs-2943	+5.63	1.5	0.3		
He	Cs-2928	+5.60	1.5	0.3		
He	Cs-2917	+5.50	1.6	0.3		
CH_4	K-4047	−1.02	1.77	1.7	13.7	Füchtbauer and Reimers[d]
	K-4044	−1.40	1.96	1.4		
C_2H_6	K-4047	−1.16	1.53	1.3	16.2	
	K-4044	−1.61	1.80	1.1		
C_3H_8	K-4047	−1.25	1.35	1.1	17.8	
	K-4044	−1.27	1.72	1.0		

[a] See reference 107.
[b] See reference 108.
[c] See reference 109.
[d] See reference 124.
[e] See reference 110.
[f] See reference 117.
[g] See reference 103.
[h] See reference 116.
[i] See reference 114.
[j] See reference 115.
[k] See reference 123.

and A. On the other hand, Hull[107] found very large T effect on the K resonance line perturbed by nitrogen, about twenty times larger than what the theories predict, at low pressures. Furthermore, Horodniczy and Jablonski observed a reduction of asymmetry with temperature, which has been confirmed by Gorodnichus[135] with Hg/A.

4. Broadening of the High Member Lines

The broadening of high member lines is found to be linearly proportional to the r.d. as is the shift, independent of the nature of absorbing atoms, but dependent on the nature of the foreign gas. The broadening of the higher series lines of K produced by He, and of Na produced by He, Ne, A, H_2, and N_2 were observed by Füchtbauer and Heesen[123] and Schulz.[136]

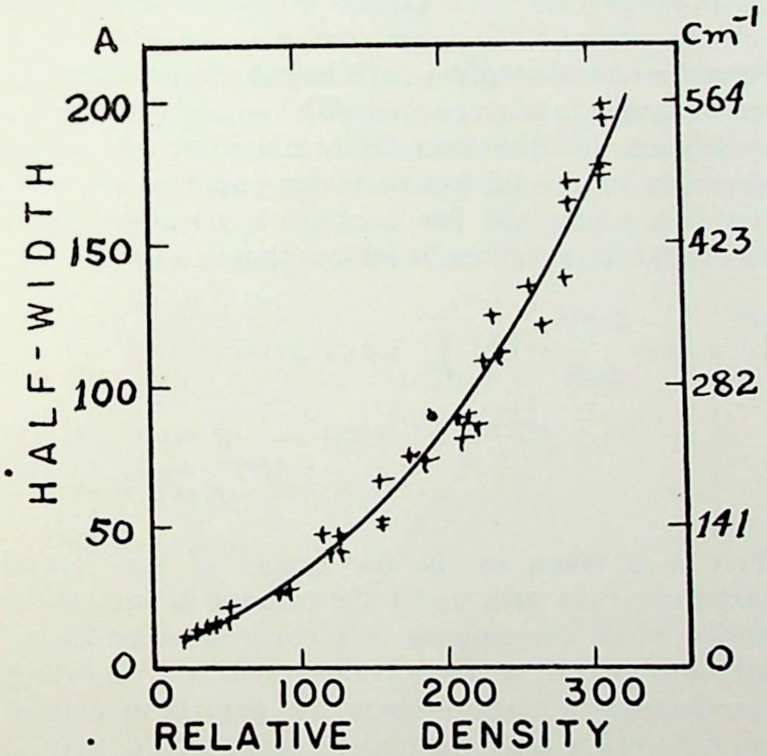


FIG. 10. Half-width of sodium D lines *vs* relative density of argon (after Robin, Robin, and Vodar[82]).

They found that the half-width *vs* series member curve had a broad maximum at the 3rd member for K/He and at the 5th member for Na/A and Na/He. Then the broadening turned back to a stationary value for high series members, *viz.* from the 11th to 12th member up for K/He (from the 9th or 10th member up for high r.d.), and from the 16th member up for Na/A, Na/He, Na/N_2, Na/Ne, and Na/H_2, as shown in Figs. 12 and 13. The broadening of the series end of the Na principal series in cm^{-1}/r.d. for He, Ne, A, Kr, Xe, H_2 and N_2 is 1.88, 1.42, 4.45, 8.12, 9.7, 6.16, and 6.16, respectively. A very large broadening of the sodium high member lines produced by Cs vapor was reported by Füchtbauer and Heimann.[126]

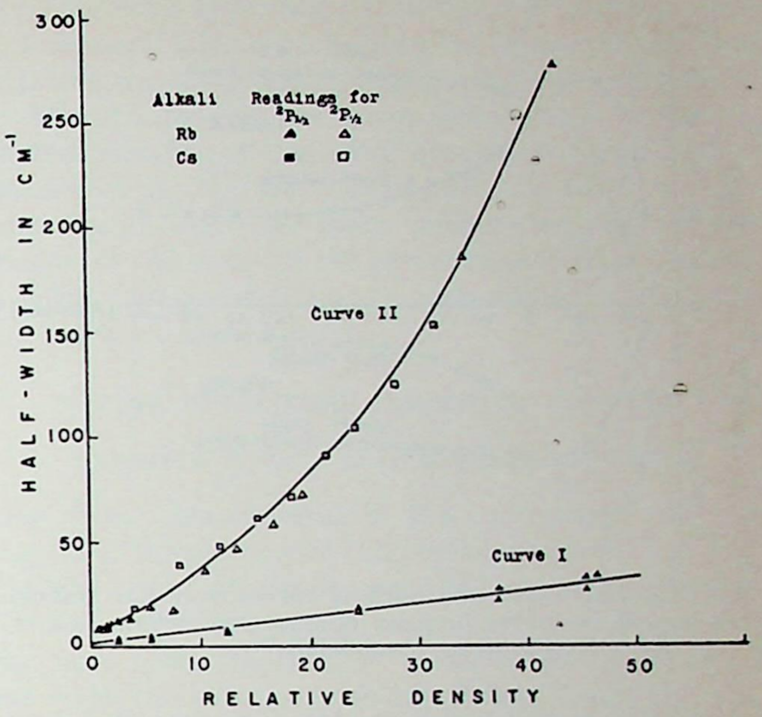


FIG. 11. Half-width *vs* relative density of He, as in Fig. 9.

The dependence of $\Delta\nu_{\frac{1}{2}}$ on r.d. for various members is shown in Fig. 14 for the case of K/He. The increase of broadening with r.d. (i.e., the slope of the curve) for the 4th and the 5th members of the K absorption series is greater for low densities than for high densities. The greater portions of the curve for the 6th and the 7th members become straight for higher densities. The readings for the 8th up to the 11th members are crowded together rather closely in some locations. The curves are linear or slightly concave upwards for high r.d.'s.

It is interesting to note in Table III that the average

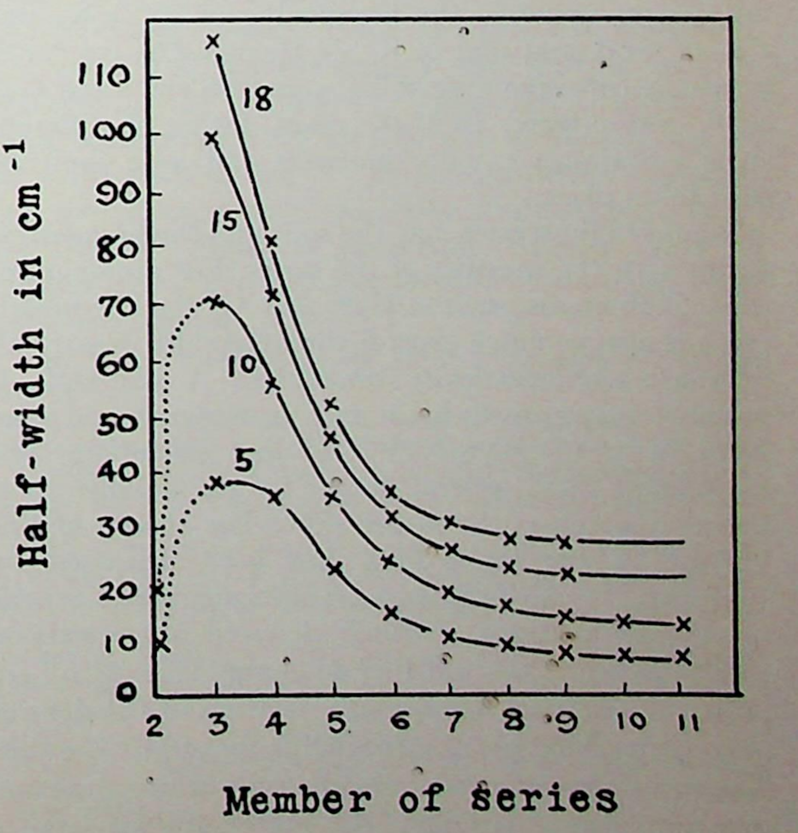


FIG. 12. Half-width *vs* members of the principal series of K for relative densities 5, 10, 15, and 18 of helium.

[135] G. A. Gorodnichus, Izvest. Akad. Nauk. S.S.S.R. Ser. Fiz. 18, 255 (1954).
[136] C. Füchtbauer and P. Schulz, Z. Physik 97, 699 (1935).

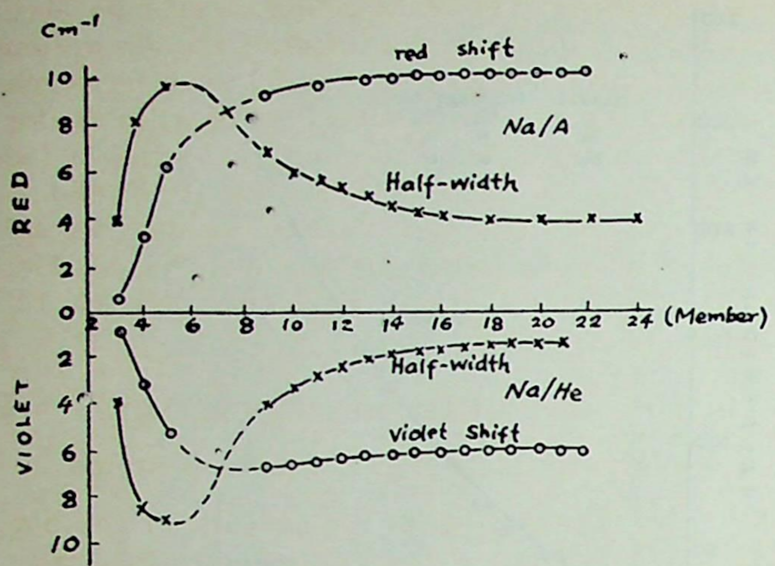


FIG. 13. Half-width and shift *vs* members of the principal series lines of Na in the presence of argon and helium. (r.d. = 1, $P = 762°K$) (after Füchtbauer and Schulz[136]).

of the ratio of half-width to shift for argon is 2.2 and that for N_2 is 2.3, which is in good agreement with Margenau's theoretical computation.[66] This ratio can be found easily also from Eqs. (97) and (98) to be 2.77, or from Eq. (58) to be 2.75, by taking $p=6$ for dispersion type interaction. Ne gave unexpectedly a rather large value of 4.4. The corresponding values for H_2, 4.1, and He, should not be included when comparison with theory is made.

Numerous cases have shown that for alkali atoms a line which undergoes greater broadening also undergoes a greater shift. But for certain lower member lines, where a switch of red shift to violet shift takes place, a small shift is usually accompanied by a large broadening. As an example, the broadening of $K(2)/N_2$ was 1.8 cm^{-1}/r.d. and that of $K(2)/H_2$ was 2.64 cm^{-1}/r.d. while the corresponding shifts were 0.93 cm^{-1} and 0.28 cm^{-1}, respectively. In these cases, forces of different signs are almost equally probable and as a limit no shift takes place.

Figure 13 illustrates how the shift and the broadening varies with the member of the series. For atoms other than alkali atoms, such as Hg[137] and Ag, the line broadening is always much greater than the shift. Nearly all lines are asymmetrically broadened. A red shift is usually accompanied by a red asymmetry and vice versa, with Cs $3876/N_2$[116] the only exception. The asymmetry is, in general, larger for the doublet component that gives the larger shift. The theory of line breadth at the series limit has been discussed by Reinsberg[3,128b] with rather good agreement with experimental observations. Firsov[129] deduced an expression for the shift and the width of an energy level with high principal quantum number n in relation to the density of the perturbing gas N surrounding the atom. The shift is shown to be independent of the type of the atom and of n provided n is large. The half-width for large n and N is proportional to $N^{\frac{1}{3}}$ and n^{-3}. As a consequence, the width of spectral lines, for very high n, does not depend on n. His theoretical curves of the line width *vs* n relationship lie lower than the experimental one because the thermal motion of foreign gases is not taken into account.

5. The Evaluation of Some Physical Quantities

As described in Margenau and Watson's previous review article[2] the oscillator strengths f and transition probabilities A can be computed from the measurement of the amount of total absorption, $\int_0^\infty (n_k)d\nu$, by graphical integration of the line contours. n_k is the so-called index of absorption defined by the relation

$$n_k = \frac{\lambda}{4\pi l}\log i_0/i, \qquad (137)$$

where l is the absorption path length. In practice, the line contours are often plotted with $\log_{10} i_0/i$ (or density) as ordinate and the distance in mm from the central maximum of the microphotometer curve as abscissa. The area under the line contour is measured by a planimeter in cm^2. Then it follows that:

$$\int_0^\infty n_k d\nu = \frac{2.303}{4\pi l\lambda} 3\cdot 10^{10} \int_0^\infty \log_{10}(i_0/i)\delta\lambda$$
$$= \frac{2.303}{4\pi l\lambda} 3\cdot 10^{10} \frac{B\cdot 10^{-8}}{GMD} a, \qquad (138)$$

where λ is taken as the wavelength of the central maximum, a the area under the contour in cm^2, D the density which corresponds to 1 cm in abscissa of the plot, and B/GM is the wavelength $\delta\lambda$ (in A) which corresponds to 1 cm in ordinate, M being the magnification ratio of the microphotometer trace, and G that on the graph.

According to radiation theory the oscillator strength of an atom is

$$f = \frac{4\nu m_a}{Ne^2}\int_0^\infty n_k d\nu = \frac{108\cdot 10^2}{22.75}\cdot\frac{1}{\lambda N}\int_0^\infty n_k d\nu, \qquad (139)$$

for $4\pi l\pi_k/\lambda \ll 1$ and the transition probability per unit time, A,

$$A = \frac{4\pi}{cNh}\int_0^\infty n_k d\nu = \frac{\pi e^2}{ch\nu m_e} f. \qquad (140)$$

From such measurements Ch'en[110] computed the f values of the Rb resonance lines under different pressures of foreign gases up to r.d. 45.52 for helium and 46.44 for argon. The f values without the pressure of foreign gases were found by extrapolating to zero density of foreign gas. The results (0.33 and 0.66 for the $^2P_{\frac{1}{2}}$ and $^2P_{\frac{3}{2}}$ components, respectively) are in agreement

[137] P. Schulz, Z. tech. Phys. 19, 585 (1938).

with Füchtbauer and Hofmann's determination[138] that the total oscillator strength is nearly unity. Although the precision of determination was not particularly high, it seems that the total absorption decreases with increasing r.d., so that f values decrease when the line is broadened. This was also true with Hg.[111] By means of exactly the same method, Michels and Kluiver[86b] obtained the oscillator strength of Hg 2537 as 0.0255 from their observations with He.

The transition probabilities, 6.93×10^9 and 3.53×10^9 for the Rb $^2P_{\frac{1}{2}}$ and the $^2P_{\frac{3}{2}}$ components, respectively, are in good accord with the results of the resonance lines of other alkalis.

The optical collision diameters can be computed from the observed half-widths, using the relation

$$\rho^2=[\pi\Delta\nu_{\frac{1}{2}}/2N(2\pi kT)^{\frac{1}{2}}]\cdot[mM/(m+M)]^{\frac{1}{2}}, \quad (141)$$

where m and M are the masses of the absorbing and of the perturbing atoms, respectively. The accuracy of the calculation depends greatly on that of the values of half-widths per unit r.d. As the slopes of these curves change with r.d. and are different for different doublet components, it is hard to assign a proper value. Although the average values of the 4th column of Table III were used, the figures listed in the 6th column can not be regarded as accurate. In spite of these limitations, the values of ρ for a given member of the absorption series and a given foreign gas are found to be very close for different alkalis. The mean values of ρ (in A) for the first member of the alkali absorption series in the presence of A, N_2, He, and H_2 are 13, 10, 8, and 7, respectively; and the mean values for the second member of the series in the presence of A, N_2, Ne, He, H_2, CH_4, C_2H_6, and C_3H_8 are, respectively, 21, 16, 12, 12, 12, 14, 16, and 18 A.

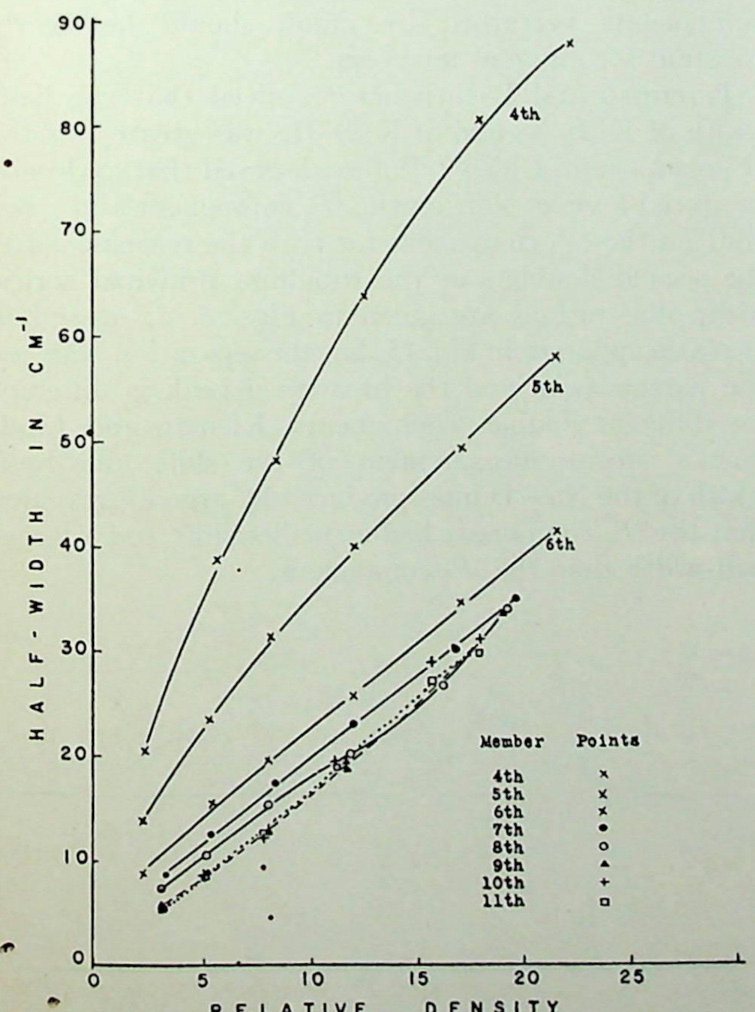


FIG. 14. Relationship of half-width and the relative densities of helium for high members of K principal series (after Füchtbauer and Heesen[123]).

B. FINE STRUCTURE PRESSURE EFFECTS

1. Pressure Effects Due to Identical Atoms

By direct measurement of the contours of the line wings, the broadening of the resonance lines of nearly all alkali atoms in absorption under various pressures of their own homogeneous vapor (resonance broadening) have been observed.[139–142] All observers reported that when the pressure was low (below 1 mm Hg) the line contours could be described nicely by the dispersion formula and the broadening was essentially symmetrical. The half-width varied linearly with the concentration of the vapor. All but the first, who used K, found that the half-width of the transition from the $^2P_{\frac{3}{2}}$ level was greater than that from $^2P_{\frac{1}{2}}$ level by a factor of about $\sqrt{2}$, quite in harmony with Foley's theoretical result[22] [see Eq. (131)]. Not only the ratios, but also the absolute values of half-widths given by Eq. (131), are found to be in accord with observations. However, when the thermal energy per perturber is comparable with the doublet separation, the theory will undergo some revision due to the fact that the J representation breaks down in such a case.[143]

Ch'en[140] and Gregory[142] observed also that indications of a violet and red asymmetry were present at high pressures for the $^2P_{\frac{3}{2}}$ and $^2P_{\frac{1}{2}}$ components, respectively, and a definite band on the red side of the $^2P_{\frac{3}{2}}$ and on the violet side of the $^2P_{\frac{1}{2}}$ component were observed. This strong asymmetry was also observed with Xe[144] and Hg.[76,77] Tomiser's observation shows that the broadening of Zeeman components are not a linear function of density.

This self-broadening has also been observed in a

[138] C. Füchtbauer and W. Hofmann, Ann. Physik **43**, 96 (1914).

[139] K: P. E. Lloyd and D. S. Hughes, Phys. Rev. **52**, 1215 (1937), at pressures up to 20 mm Hg. $\Delta\nu_{\frac{1}{2}}(J=\frac{3}{2})/\Delta\nu_{\frac{1}{2}}(J=\frac{1}{2})=1.0$. J. Tomiser, Acta Phys. Austriaca **8**, 325 (1954), observed up to 0.1 mm Hg with the result of ratio, 1.07–1.59 (mean: 1.41) (theor. 1.42).

[140] Rb: S. Y. Ch'en, Phys. Rev. **58**, 844 (1940), at pressures up to 152 mm Hg. $\Delta\nu_{\frac{1}{2}}(J=\frac{3}{2})/\Delta\nu_{\frac{1}{2}}(J=\frac{1}{2})=1.75$. J. Tomiser, reference 139, 9, 18 (1954), up to 0.08 mm Hg with ratio 1.19–1.58 (mean: 1.26) (theor. 1.38).

[141] Na: K. Watanabe, Phys. Rev. **5?**, 151 (1941), at pressures up to 70 mm Hg. $\Delta\nu_{\frac{1}{2}}(J=\frac{3}{2})/\Delta\nu_{\frac{1}{2}}(J=\frac{1}{2})=1.55$. J. Tomiser, reference 139, 8, 198 and 276 (1954), up to ü.9 mm Hg with ratio 1.27–1.45 (mean 1.31) (theor. 1.42).

[142] Cs: C. Gregory, Phys. Rev. **61**, 465 (1942) at pressures up to 17.5 mm Hg. $\Delta\nu_{\frac{1}{2}}(J=\frac{3}{2})/\Delta\nu_{\frac{1}{2}}(J=\frac{1}{2})=1.72$ (theor. 1.52).

[143] M. Takeo, Bull. Am. Phys. Soc. Ser. II, **1**, 245 (1956).

[144] J. C. Madenann and R. Turnbull, Proc. Roy. Soc. (London) **A129**, 265 (1930); **A139**, 683 (1933).

discharge tube of mercury vapor on the lines Hg 10140 A ($2^1P_1-2^1S_0$) and Hg 5770 A ($2^1P_1-3^3D_2$). The broadening of the former line is purely due to self-pressure, while the latter was broadened proportionally to the electric current.[105] Thus, by extrapolating the latter broadening to zero current, the self-broadening of the line was obtained for each corresponding pressure.[145] Thus, the two lines were found to behave according to the impact theory with the same width. However, the experimental absolute width was 60% larger than the theoretical value.

2. Pressure Effects Due to Foreign Atoms

The difference in behavior of shift and broadening of the doublet components of the principal series lines of alkalis and of Ag was first emphasized by Ch'en and his collaborators.[110,112,113,117] For example, the shift of the first doublet of the Rb principal series produced by helium[110] was -6.88 A ($+10.88$ cm^{-1}) for the $^2P_{1/2}$ component and -2.68 A ($+4.40$ cm^{-1}) for the $^2P_{3/2}$ component for r.d. 45.52 as determined by a 21-foot grating. The corresponding figures for the half-widths were 16.9 A (27.8 cm^{-1}) and 20.2 A (33.2 cm^{-1}), respectively. The ratio of shifts for the two components is 2.5 and that of half-width is 0.84.

The shift of the Rb second doublet[113] produced by 43.9 r.d. of helium was -19.5 A ($+110$ cm^{-1}) for the $^2P_{1/2}$ component and -12.2 A ($+69.0$ cm^{-1}) for the $^2P_{3/2}$ component. The corresponding values for the half-width were 48.7 A (277 cm^{-1}) and 63.4 A (361 cm^{-1}), respectively. The ratios of shifts and half-widths are 1.60 and 0.766, respectively.

For argon, the difference of shifts of the two doublet components of all alkalis is very small, but the broadening of the $^2P_{3/2}$ component is slightly greater than that of the $^2P_{1/2}$ component by a factor of 1.12 on the average.

The line shift, as discussed in Sec. II.B.b, originates in the smaller phase shift in the oscillation of an absorber at a collision. The phase shift produced at a certain impact parameter depends strongly on the shape of the interatomic potential curves for various gases. For helium an impact parameter effective for line shift will be larger than the optical diameter, and the classical discussion of the shift may be adequate at rather high temperature by the following reason. Namely, for alkali resonance lines with helium as perturbers, the optical collision diameter happens in many cases to be nearly of the same order of magnitude as the sum of the atomic diameter of helium (3.5 A) and of the absorbing atom, both in the ground state. When the absorber is excited, the overlap integral will be appreciably larger than otherwise, and a repulsion will dominate the second-order dispersion energy at the distances for line shift impact parameters. Thus, a violet shift with helium pressure may be expected. The shift for higher members should be larger, not because of the fact that the overlap integral is larger for higher states, but rather because of the fact that the number of collisions effective in line shift increases for higher states. However, this line shift is related to the shift of the center of gravity of the doublet.

The difference of shifts for doublet components must be found in the variation of the doublet separation, which is determined by the change of the wave function at the absorber's nucleus. The larger impact parameter, effective for higher-member line shift, is unlikely to change the doublet separation more than otherwise. Thus the ratio of shifts is smaller for higher members. On the other hand, observation shows that for alkali resonance lines with argon pressures the optical collision diameter is about 1.5 times larger than the sum of atomic diameters concerned and lies well inside the region dominated by dispersion forces. In this case, the ratio of shifts for the doublet components will be constant for different members. The same kind of argument may follow for the ratio of half-widths for doublets. But the phase shift most effective for broadening would be larger than that produced at the optical collision diameter. For helium, the impact paramerer lies inside a quantum-mechanical region, although for argon the classical picture is still applicable, using the result in Sec. II.A.b, that the ratio of half-widths for doublet components perturbed by argon should be nearly constant for different members.

Peterman and Füchtbauer[124] noticed that the half-width of K(2)/N_2 and of K(2)/He was greater for the $^2P_{3/2}$ component. Ch'en[110,115] also observed that hydrogen produced a violet shift on the $^2P_{1/2}$ component and a red shift on the $^2P_{3/2}$ component for both the resonance and the second doublets of the rubidium principal series. More illustrations are given in Figs. 3, 5, 6, and 9. It is also apparent in Fig. 15 that the separation between the narrow peak and the broadened peak is different for different doublet components. Kleman and Lindholm's precise measurement of the shift and half-width of the Na—D lines produced by argon[107] revealed that the $^2P_{1/2}$ component had a smaller shift and a larger half-width than the $^2P_{3/2}$ component.

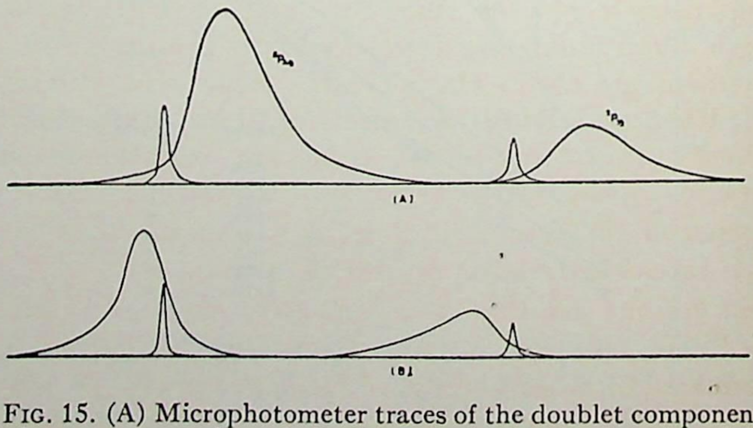


FIG. 15. (A) Microphotometer traces of the doublet components of the second member of the Rb principal series broadened by argon at pressures 0.1 atmos (narrow peaks) and 12.18 atmos (broad peaks). (B) Corresponding traces for helium at pressures of 0.12 and 10.92 atmos, respectively. Note the difference in shift and broadening of the two components (after Ny and Ch'en[117]).

[145] P. Gerthsen, Z. Physik 138, 515 (1954).

Recent observations by Robin and Robin[85] showed that the shift of the $^2P_{\frac{1}{2}}$ component of Rb resonance lines when perturbed by A, He, and N_2 up to several hundred r.d. was nearly linear with r.d., while that of the $^2P_{\frac{3}{2}}$ component reversed its direction of shift for r.d. equal to 180 for N_2 and 260 for A. Because of the fact that the $^2P_{\frac{3}{2}}$ component was broadened more rapidly than the $^2P_{\frac{1}{2}}$ component, the relative intensities of the two lines varied very much with pressure. The less intense $^2P_{\frac{1}{2}}$ component at low pressure became more intense at high pressures.

From the results of observations obtained so far one can make the following general empirical statements:

(1) The difference in shift (when expressed in cm^{-1}) of the multiplet components depends on the member of the series and the nature of foreign gases, but is independent of the nature of the absorbing atom.

(2) For gases that produce a violet shift, the longer wavelength component ($^2P_{\frac{1}{2}}$) exhibits a greater violet shift than the shorter wavelength component ($^2P_{\frac{3}{2}}$), as predicted at the end of Sec. II.C.b, and the ratio for the first doublet is larger than that for the second doublet.

These two statements are supported by all available data, such as Rb(1)/He, Rb(2)/He, Cs(2)/He, and Ag(1)/He without exception. For Rb(2)/Ne, the $^2P_{\frac{1}{2}}$ component produced a violet shift while the $^2P_{\frac{3}{2}}$ component produced a red shift. For a given r.d. of He, the differences of shift (in cm^{-1}) of the doublet components for the second member of Rb and Cs principal series are the same within experimental error.

(3) The difference in shift of the double components for gases that produce a red shift is usually small as shown in Sec. II.C.b. There is no apparent indication as to which component should show a stronger red shift. For Rb(1)/*A*, Rb(2)/A, Cs(2)/A, K(1)/A, Ag(1)/A, K(1)/N_2, and K(2)/N_2 the $^2P_{\frac{1}{2}}$ component shows a slightly stronger red shift; while for Na(1)/A Rb(2)/N_2, and Robin's high pressure observations of Rb(1)/A and Rb(1)/N_2 the $^2P_{\frac{3}{2}}$ component has a slightly stronger red shift.

(4) The broadening of the short wavelength ($^2P_{\frac{3}{2}}$) component is greater than that of the longer wavelength ($^2P_{\frac{1}{2}}$) component. Observations confirming this statement are Cs(2)/H_2, K(2)/H_2, Rb(1)/He, Rb(2)/He, Rb(1)/A, Rb(2)/A, and Ag(1)/A. Robin and Robin[85] proved indirectly, from the variation with pressure of the ratio of the optical densities of the maxima of the resonance lines of K and Rb, that the $^2P_{\frac{3}{2}}$ is broadened more than the $^2P_{\frac{1}{2}}$ component.

Exceptions to this are Na(1)/A, Rb(2)/Ne, and Rb(2)/He only when the r.d. is low. As shown in reference 113, when the r.d. of He is above 19, the half-width of the short wavelength component becomes greater than that of the longer wavelength one.

It is to be noted that in this and the following sections there is demonstrated very strikingly the important fact that there exists a conspicuous difference in the interactions at close distances for the two fine structure levels. There is no doubt that this phenomenon will be useful in interpreting certain detailed phenomena of various basic observations.

Many research problems dealing with these effects can be suggested, such as the pressure effect on high multiplicity components, the fine-structure pressure effects of multiplet components in relation to different couplings of the angular momenta of the optical electrons, etc.

C. NARROW DIFFUSE BANDS OF VARIOUS METALS PRODUCED BY CLOSE ENCOUNTERS WITH FOREIGN PARTICLES

The most intense part of a broadened spectra line should come from the majority of the perturbers and corresponds to distant collisions. The shapes of the wings, however, will depend on the phenomena occurring close to the absorber where spectral lines are affected most strongly. This phenomenon is manifested as narrow diffuse bands which appear as faint wings of the atomic lines if they are not resolved. They have been observed very conspicuously as separate bands or diffuse lines in the neighborhood of certain atomic lines. These bands which are often called "satellites," are present at foreign gas pressures even as low as a few milimeters of Hg.

1. Observations in Emission Spectra

This phenomenon was probably first observed by Oldenberg,[146] who obtained a narrow continuum of this sort on both sides of Hg 2536.7 in fluorescence by the irradiation of the mercury resonance line in a mixture of mercury and a few cm pressure of helium. When heavier rare gases were used one or two diffuse intensity maxima were found on the short wavelength side, lying closer to the atomic line the greater the atomic weight of the rare gas used. Kuhn and Oldenberg[147] interpreted the bands as due to loosely bound molecules or space quantization.

Krefft and Rompe[148] observed similar diffuse bands in the positive column of electric discharges in mixtures of Tl, In, Zn, Na, K, Rb, and rare gases. The vapor pressure of the metal was about 1 mm and the gas pressure about 2–10 mm. They found that the characteristics of the bands depend upon the rare gas used and also upon the atomic levels of the metal. These bands were classified according to visual judgment of their shape, and a number of empirical relations were suggested, such as the proportionality of the separation of certain violet bands to $1/\sqrt{M}$, M being the atomic weight of the rare gas used.

Preston[149] made an extensive study of these bands in

[146] O. Oldenberg, Z. Physik 47, 184 (1928); 55, 1 (1929).
[147] H. Kuhn and O. Oldenberg, Phys. Rev. 41, 72 (1932).
[148] H. Kreft and R. Rompe, Z. Physik 73, 681 (1932).
[149] W. M. Preston, Phys. Rev. 51, 298 (1937).

a simple capillary discharge tube of Hg, Cd, or Tl in the presence of 10 cm of He or A. He used a Hilger medium quartz spectrograph, and found that the intensity of the band was of the order of a few hundredths to two thousandths of the intensity of the accompanying metal line. The Hg 2537 band at 2526.0 A was not changed within a limit of error of about 0.3 A for three different temperatures, 55, 130, and 400°C. Also, no observable shift was reported when argon pressure was varied from 1 mm to 10 cm, although the band may be visible, with much less intensity, even with 0.1 mm of argon. He identified the collision-induced spectra near 13 lines of the spectrum of Hg, 12 of Cd, and 6 of Tl. The majority of these spectra had roughly defined edges at the shorter wavelength side and no resolved maximum. Only three definite cases of maxima on the short wavelength side of lines were encountered, all lying fairly close to and definitely associated with an atomic line. Several examples were also found of maxima at varying distances on the long wavelength side of lines, some at such great distances that they could not be definitely connected with any particular line. He pointed out that the bands associated with the lines of a particular spectral multiplet are much more alike in form and width than those belonging to unrelated lines.

It should be noted that the conditions for the appearance of the bands in emission, such as the presence of ions, electron impacts, inhomogeneity in temperature distribution and in concentration of emitting atoms, etc. in the source, could hardly be neatly defined for theoretical analysis.

2. Observations in Absorption Spectra

a. Observations with Low Foreign Gas Pressures

Although the appearance of a new band of Hg at 2528 A due to the presence of foreign gases was first known to Moore[150] and Oldenberg[146] systematic observations of the structureless alkali-foreign-gas absorption bands were perhaps first started by Ch'en and collaborators.[151–153,78,79] With the Bausch and Lomb or the Hilger large Littrow spectrograph the band appearing on the short wavelength side of the second member of the principal series of rubidium in the presence of foreign gases (H_2, He, N_2, and Ne) was first observed, and then the observation was extended to the second member of cesium and potassium, and the third member of the sodium principal series.[152] In the presence of helium, bands were observed near the fourth through the seventh member of the lithium principal series.[78] Recently, with infrared plates and more careful technique, the bands appearing in the neighborhood of the resonance lines and the third or higher members of the rubidium principal series were also observed with a number of other gases,[79] *viz.*, A, Kr, and Xe. Most recently Ch'en and Jefimenko[153] first confirmed that the narrow bands exist simultaneously both on the red and on the violet side of the first few principal series lines of Rb and Cs, with the red bands having a higher intensity. In addition to the well-defined series of rare gases (also N_2 and H_2), fourteen different kinds of paraffin molecules, whose physical properties are very well known, were employed as a series of perturbing particles. Over two hundred more new bands of this kind were obtained.[153] For details one should refer to these original papers.

Only with heavy foreign gases, such as xenon and krypton, and saturated hydrocarbon vapors heavier than propane, was the narrow diffuse band resolved on the *red side* of each component of the absorption doublets of Na, Rb or Cs. Figure 16 is a set of sample spectra for the Rb/Xe and Cs/Xe "red" bands. The red

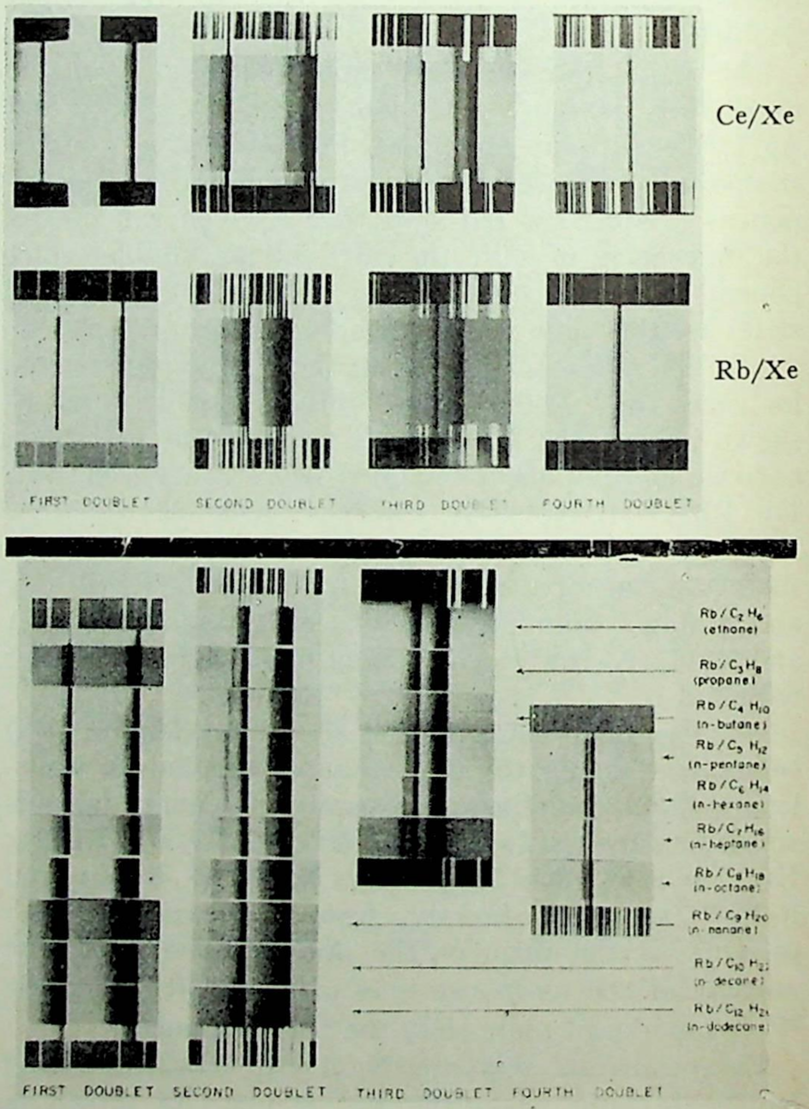


FIG. 16. (a) The red bands of Rb and Cs in the presence of xenon (1) for the resonance lines, (2) for the second doublet, (3) for the third doublet, and (4) for the fourth doublet of the principal series. The corresponding pictures for the fifth doublet showed only some red asymmetry (after Ch'en and Jefimenko[153a,b]). (b) The red bands of Rb in the presence of various paraffin vapors (after Ch'en and Jefimenko[153b]).

---

[150] H. R. Moore, Science 66, 543 (1927).
[151] T. Z. Ny and S. Y. Ch'en, Nature 138, 1055 (1936); J. phys. 9, 169 (1938).
[152] S. Y. Ch'en, Phys. Rev. 65, 338 (1944).
[153] (a) S. Y. Ch'en and O. Jefimenko, J. Chem. Phys. 26, 256 (1957) (results for Rb). (b) O. Jefimenko and S. Y. Ch'en, J. Chem. Phys. (to be published) (Cs).

TABLE IV. The $\Delta\nu_m$ (in cm^{-1}) values for alkali-foreign gas red bands for the various members of the principal series.

	$\Delta\nu_m$ (1st member)		$\Delta\nu_m$ (2nd member)		$\Delta\nu_m$ (3rd member)		$\Delta\nu_m$ (4th member)	
	$^2P_{1/2}$	$^2P_{3/2}$	$^2P_{1/2}$	$^2P_{3/2}$	$^2P_{1/2}$	$^2P_{3/2}$	$^2P_{1/2}$	$^2P_{3/2}$
Na/Xe	-26.6[a] cm^{-1}	...	-76 cm^{-1}	...	-47 cm^{-1}	...	-16 cm^{-1}	...
Rb/Xe	-17.4	-33.3	-45.2	-43.5	-34.4	-21.0	-10.7	-9.5
Cs/Xe	-11.7	-31.4	-37.7	-52.7	-29.3	-29.4	-6.1	-9.4
Rb/*n*-octane	-64[a]	-76[a]	-53.0	-46.5	-21.9	-20.9	-10.5	-10.6
Cs/*n*-octane	-47[a]	-69[a]	-46.7	-49	-19.8	-19.9	-8.2	-9.7

[a] The accuracy determination is low, because the absorption maxima could not be definitely established or resolved from the atomic line.

bands become perceptible when the vapor pressure of Rb or Cs is of the order of 10^{-3} mm Hg or higher (temperature of absorption tube $=120-150$°C) and the pressure of foreign gas is higher than 20 cm Hg.¶¶ When the vapor pressure is increased (by increasing the temperature of the absorption tube to 200–270°C) the red bands and the absorption lines of Rb or Cs overlap, and the bands are no longer resolved from the broadened absorption line. Under these conditions the violet bands become intense enough for observation. When the temperature of the absorption tube is still increased (400–500°C) a number of additional bands are observed between the doublet components of the resonance lines. Owing to the fact that the appearance and the regularities of occurrence of these types of bands are quite different. They are described in separate sections.

(*i*) *The Red Bands.*—As shown in Fig. 16 the red bands appear in general close to the atomic line as a single absorption maximum. No red bands are produced by light gases, H_2, D_2, He, Ne, and N_2. Argon and methane merely give a red asymmetry of the atomic lines. A definitely resolved red band is observed for each component of the first four doublets of the principal series of Na, Rb, and Cs for all the other heavier rare gases and saturated hydrocarbons employed.

Observations connected with the *red bands* associated with the principal series lines of alkali atoms may be summarized as follows: (1) The separation (in cm^{-1}) between the band and the associated absorption line, $\Delta\nu_m$, varies strongly with the member of the principal series. For Kr and Xe, the largest values of $\Delta\nu_m$ are those associated with the second member of the series, while for all paraffin molecules the highest value of $\Delta\nu_m$ falls on the first member, and $\Delta\nu_m$ decreases rapidly with the increase in the ordinal number of the member of the series, as illustrated in Table IV. No bands were resolved from the atomic lines of fifth and higher doublets of the series. (2) The values of $\Delta\nu_m$ are different for the two doublet components. For both rare gases and saturated hydrocarbons all values of $\Delta\nu_m$ associated with the short wavelength ($^2P_{3/2}$) component of the resonance lines are considerably greater than that associated with the long wavelength ($^2P_{1/2}$) component. For the second member of the Rb principal series the foregoing statement holds for Kr, and light hydrocarbons (propane and butane), but is reversed for Xe, butane, and all heavier paraffin molecules. The difference becomes small for the third and the fourth member of the Rb principal series. (3) Red bands are resolved only for heavy gases. As shown in Figs. 17 (a) and (b) for any given doublet, $\Delta\nu_m$ increases linearly with the polarizability, α, of the gases (with possibly a saturation for very large particles). The slopes of the curves are almost the same for both Rb and Cs, *viz.*, 0.31 and 0.27 in cm^{-1}/α (α being expressed in 10^{25} cm^3) for the $^2P_{1/2}$ and the $^2P_{3/2}$ components of the second doublet, respectively. The corresponding values for the third doublet are both 0.11. (4) Although there are no sufficient data, the values of $\Delta\nu_m$ for $^2P_{1/2}$ component for a given foreign gas appears to be a little higher for lighter alkalis. (5) The positions of the maxima of the bands is not appreciably changed by either a variation of pressure from $\frac{1}{2}$ to 2 atmospheres, or a change of temperature from 120° to 190°C. (6) The widths of bands decrease with the ordinal member of the doublets and are roughly of the same order of magnitude as those of the associated

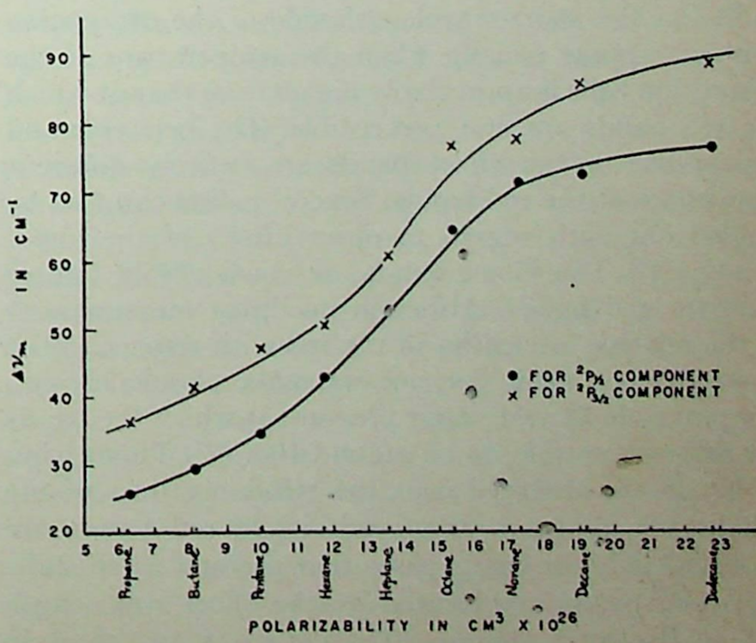


FIG. 17. (a) $\Delta\nu_m$ *vs* the polarizability of paraffin molecules for the 1st member of the Rb principal series (after Ch'en and Jefimenko[153a]).

¶¶ If the foreign gas pressure is much lower than this value, the vapor pressure of Rb must be increased to make the band intense enough for observation.

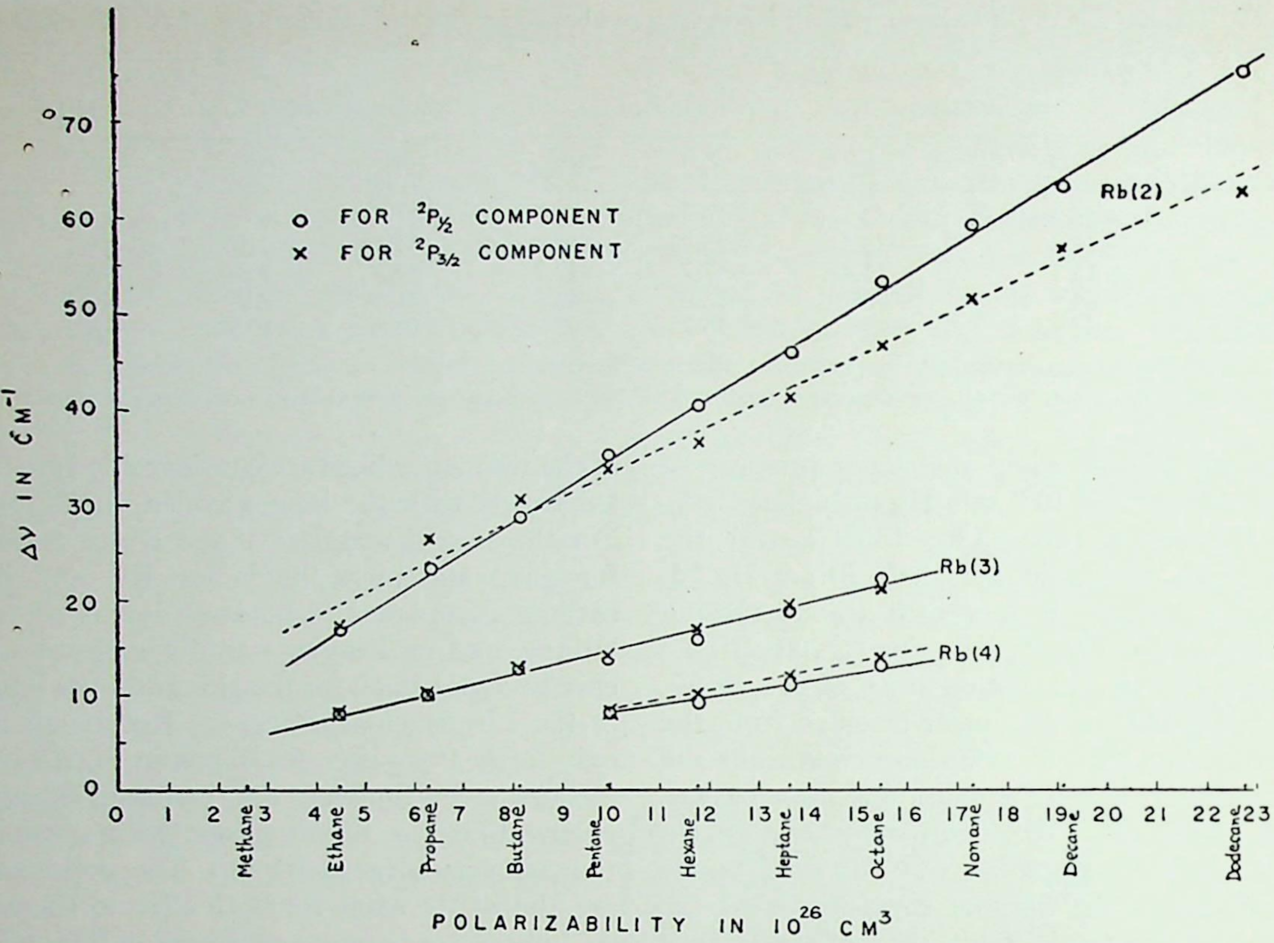


FIG. 17. (b) $\Delta\nu_Q$ *vs* the polarizability of paraffin molecules for the second, the third and the fourth members of the Rb principal series (after Ch'en and Jefimenko[153a,b]).

absorption line. (7) The intensity of the bands relative to that of the associated atomic line is higher for bands with smaller $\Delta\nu_m$ or smaller widths. When the pressure of foreign gas is low, say below 3 cm of Hg, the intensity depends strongly on both the concentration of the absorbing atom and that of the foreign gas. (8) The red bands are observed only for those gases which produce a red shift in pressure effects.

(*ii*) *The Violet Bands.*—The violet bands, i.e., the bands on the short wavelength side of the absorption doublet, appear usually when the temperature of the absorption tube is appreciably higher than that at which the red bands are first perceptible. The behavior and appearance of the violet bands are entirely different from those of the red bands. Several points can also be singled out with regard to observations of the violet bands: (1′) The violet bands are considerably fainter than the red bands. Although no direct measurement of the relative intensities of the red with respect to the violet bands is made yet, one can make a rough estimation, from the lowest vapor pressure at which the bands are first perceptible, to be around 100. (2′) Violet band or bands are observed near the resonance lines of the alkalis for all gases employed, while red bands are observed only for heavy gases that produce a red shift. (3′) The bands are located at the short wavelength side of the entire doublet. The values of $\Delta\nu_m$ given in Table V are measurements from the short wavelength component ($^2P_{\frac{1}{2}}$) of the doublets. (4′) Several maxima may be observed near the resonance lines. The values of $\Delta\nu_m$ for the violet bands near the resonance lines are very much greater than those of the red bands, but decrease much more rapidly towards higher members of the principal series than in the case of the red bands. For Rb, the values of $\Delta\nu_m$ for the second member of the series are decreased by 8, 12, and 20-fold relative to those for the first member, for He, Ne, and A, respectively. If one followed this trend of decrease, the values of $\Delta\nu_m$ for the second member would be 11 and 8 cm⁻¹ for Kr and Xe, respectively, so, experimentally, the band is never resolved. For the third member of the Rb absorption series, the violet bands are never observed

TABLE V. The $\Delta\nu_m$ (in cm⁻¹) values for various alkali-foreign gas violet bands.

Alkali \ Foreign gas	H_2	He	N_2	Ne	A	Kr	Xe
Na(1)	1203	1200	1201	1086	1069	906	893
Na(2)	...	...	...	136	50	...	...
Na(3)	52	50	55	53	...	...	...
K(2)	230	174	174	62	...	...	...
Rb(1)	921	778	718	448	433	342	334
Rb(2)	133	98	112	38	22	...	...
Rb(3)	...	...	18	...	...	...	...
Cs(1)	428	356	408	231	213	151	134
Cs(2)	78	64	60	29	...	...	...
Cs(3)	...	...	12	...	...	...	...

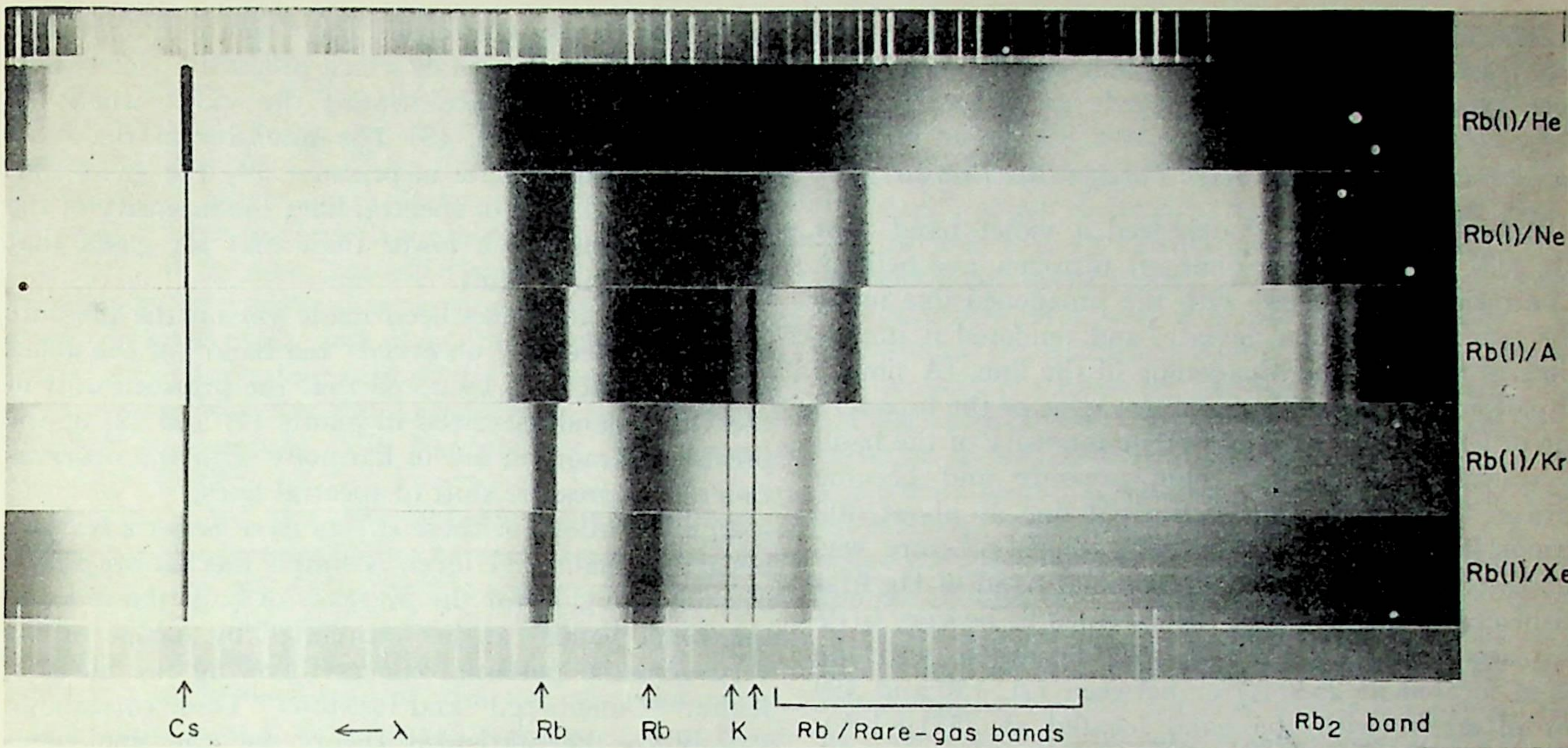


FIG. 18. The violet bands of Rb resonance lines in the presence of rare gases. The lines marked with K and Cs are the respective potassium resonance lines and the shorter wavelength component of cesium resonance lines which appeared as impurities (after Ch'en and Jefimenko[153a]).

for heavy gases; for light gases they are so close to the atomic line that only violet asymmetry was observed. (5′) As shown in Table V, the values of $\Delta\nu_m$ for violet bands (in contrast to red bands) decrease with increases in the mass of the absorbing atom. As an illustration, for a given foreign gas the $\Delta\nu_m$ values for Na(1) range from 1.5 to 2.6 times greater than those for Rb(1); the $\Delta\nu_m$ values for Rb(1) are again twice those for Cs. (6′) The most profound difference between the violet and the red bands manifests in the dependence of $\Delta\nu_m$ on the type of colliding particles. In contrast to the red bands, the $\Delta\nu_m$ values decrease with increasing mass of foreign gas particles, or remain essentially constant for paraffin series ones. This point is further detailed in next paragraph. (7′) The width of the bands is greater for those with a greater $\Delta\nu_m$. (8′) The position of the bands is not appreciably changed by a smaller variation of pressure from $\frac{1}{2}$ to 10 atmospheres, and point (7) for red bands is also true for violet bands. The effect at very high pressures for (8′) is discussed in the subsequent subsection.

As shown in Fig. 18, for a given absorbing alkali the violet bands produced by various rare gases near the resonance lines can be classified into three groups according to the magnitude of $\Delta\nu_m$: (a_1) the alkali (1)/He band which is very broad and stretched very far from the resonance lines, (b_1) alkali(1)/Ne and alkali (1)/A bands with two similar maxima: $\Delta\nu_{m1}$*** is much closer to the atomic line concerned than that for the case of (a_1) and $\Delta\nu_{m2}$ lies between them, and (c_1) alkali (1)/Kr and alkali (1)/Xe bands, in the same manner, with two maxima. $\Delta\nu_{m1}$ is still smaller than that for case (b_1). This effect may be correlated with the electron density at the surface of a perturber. The electrons having nearly the same energy associated with the atomic surface are s^2 for He, s^2p^6 for Ne and A, and $s^2p^6d^{10}$ for Kr and Xe. Since their atomic radii are nearly the same, Kr and Xe are most unlikely affected at a close encounter by whether the alkali atom is in the ground state or in the excited state.

Likewise, the violet alkali-paraffin bands near the resonance lines can be divided into three groups: (a_2) the bands in the presence of methane. They are characterized by a $\Delta\nu_m$ smaller than that for any other hydrocarbons studied, and by $\Delta\nu_{m2}$ smaller than $\Delta\nu_{m1}$. The bands resemble very closely those of (b_1)[79] in their appearance and positions. (b_2) the bands in the presence of ethane and propane. These bands have only one maximum with a $\Delta\nu_{m1}$ of apparently the same magnitude for both hydrocarbons. (c_2) the bands due to butane and heavier paraffin series. These always have two maxima, with essentially the same $\Delta\nu_{m1}$. They decrease slightly with the increase in the molecular weight of hydrocarbons; however, $\Delta\nu_{m2}$ is greater than $\Delta\nu_{m1}$ and increases slowly with the increase in the weight of the hydrocarbon. These bands resemble in appearance the violet bands in the presence of N_2.[79]

A test of the dependence of the values of $\Delta\nu_m$ on the mass of the foreign gas could only be made by employing H_2 and D_2. The positions and the appearance of the Cs/H_2 and Cs/D_2 bands are identical, within experimental errors, although the latter bands are somewhat slightly sharper.[153b]

*** The subscript "1" in $\Delta\nu_m$ refers to the value of separation for the intense maximum in the band; while subscript "2" to the much fainter maximum in the same band.

The appearance of new bands near the absorption lines of alkali atoms due to collisions with its own vapor were also observed.[78,151,154]

b. Observations with High Foreign Gas Pressures

Robin and Robin[84,86] observed a violet band near Hg 2537 in the presence of argon, nitrogen, and helium. The fusion of the band with the broadened line made the points of maxima deviate, and rendered it impossible to measure the broadening of the line. (A similar situation occurred with the observation of the broadening of Rb 4201 by helium.[103]) The intensity of the band increased rapidly with argon pressure and became greater than that of the principal line at about 300 atmospheres. As the argon or nitrogen pressure was increased, the maximum of the violet band of Hg first displaced toward the red, then seemed to be very little displaced between r.d. 250 and 400 for argon (with the band located at 2530 A) or between r.d. 150 and 350 for nitrogen (with the band located at 2531±1 A), and then displaced readily toward the violet. The authors regard this change of the maximum position as due to the change in polarizability of both absorbers and perturbers due to pressure. It has been pointed out[155] that for spherical nonpolar molecules the polarizability increases first with pressure up to about several hundreds atmospheres and then decreases. As the pressure was increased to 5000 kg/cm^2 with N_2 and 6000 kg/cm^2 with argon (temperature 95°–160°C) a second satellite band was observed by Robin and Vodar[156] at the short wavelength side of the first satellite. For helium as the pressure was increased the violet band (which appeared at 250 kg/cm^2 and was located at 2522±1 A at 600 kg/cm^2) displaced toward the violet (2516±1.5 A at 1400 kg/cm^2).

A faint (broad) violet band was also observed for Na at 5565±10 A[72] when the pressure of argon was very high (at 1000 atmos and 400°C). The Na(1)/A band observed by Ch'en and Stauffer at low pressure was at 5543.5±5 A, showing that the band was shifted towards red at high pressures. Also, the Rb(1)/A band was reported[85] as displaced toward the red (7575±10 A to 7630±20 A) up to r.d. 120, then returned toward blue (7480±30 A at r.d. 390). The Rb(1)/He band at r.d. 100, displaced toward the blue (it was at 7250±30 A at r.d. 300, 260°C). Two K(1)/A bands were observed (7370 and 7670±30 A) at r.d. 300 and 280°C, the last being only visible at high density.

So far the effect of high foreign gas pressure on the bands which appear on the violet side of the atomic line can be outlined in four points: (1) For gases which produce a red shift of the atomic line, the band will displace first toward the red and then will gradually turn back to displace toward the violet as the pressure is increased. (2) For gases which produce a violet shift, the band will displace toward the violet when the pressure is increased. (3) The intensity of the band increases with increase in pressure. (4) For gases that produce a red shift of spectral lines the intensity of the violet band is much lower than that for gases that produced a violet shift.

No observation has been made yet on the pressure shift of the recently observed "red band" of the alkali metals.[79,153] It is to be noted that the pressure shift of the violet band as stated in points (1) and (2) of the previous paragraph are in harmony with the observation of the pressure shift of spectral lines.

Interpretations of these effects have never advanced beyond a qualitative level. A simple mechanism which accounts roughly for the presence of both the red and the violet bands can be formulated by means of the very versatile potential curves following the ideas of Kuhn,[146] Oldenberg,[147] and Preston.[149] These correspond strictly to the statistical theory for the line edges described in Sec. II.C. The frequency of the light absorbed is given by the difference between the interatomic potentials for the upper and lower states concerned, since it is assumed that the energy of thermal motion of the system does not change during the electronic transition. Hence, the frequency of the absorbed light observed should be confined within the region defined by the minimum and the maximum appearing in the potential difference curve. The potential difference usually changes very slowly at the minimum and the maximum, so that rather pronounced bands will appear at the line edges corresponding to them. Thus, the minimum of the potential difference gives rise to the red band and the maximum to the violet band. If there are several minima and maxima in the potential difference curve related with an atomic line, several such resolved bands could be expected.

Red bands are not observed for light gases, such as He, H_2, and Ne, because their polarizabilities are too small to give rise to any striking minimum. Since the maximum of the potential difference appears at small interatomic distances and the radial distribution of perturbers falls off rapidly with decreasing distances, the intensity of the violet bands at low pressures are usually very weak. However, it increases enormously at high pressures,[84] probably because of the smaller mean interatomic distances.

The difference of the values of $\Delta\nu_m$ of the red bands for the two doublet components may correspond to the conspicuous variation of doublet separations at the intermediate (red band) distances of interaction according to the state of the absorber. The observed linear dependence of $\Delta\nu_m$ of the red bands with hydrocarbons on their polarizabilities certainly deserves a discussion. For a small region at intermediate distances, the R-dependent part of the potential difference curve may be approximated by $U\{(6/p-6)(R_0/R)^p$

[154] H. Kuhn, Z. Physik 76, 782 (1932); T. Z. Ny and W. P. Wong, Compt. rend. 202, 1428; 203, 429 and 860 (1936); D. K. Bhattacharyya and J. Murari, Indian J. Phys. 19, 20 (1945).
[155] L. Jansen and P. Mazur, Physica 21, 193 and 208 (1955).
[156] J. Robin and B. Vodar, Compt. rend. 242, 2330 (1956).

$-(p/p-6)(R_0/R)^6\}$ (as suggested by the Lennard-Jones potential function), where U is the minimum energy difference giving approximately the red band maximum $\Delta\nu_m$. p and R_0 are adjustable parameters, independent of the $\Delta\nu_m$. If R_0, the position of the energy minimum, is nearly the same for all perturbers, U will be proportional to the polarizability and $\Delta\nu_m$ will vary linearly with it. Also, the rapid decrease in $\Delta\nu_m$ of the red bands for higher member lines may be due to a rapid increase of R_0 with the total quantum number. Thus, the relative intensity of the band to that of the associated atomic line should increase with the ordinal number of members. This has not yet been experimentally verified.

The fact that the positions of the alkali-paraffin violet bands are essentially the same for all paraffins ($\Delta\nu_m(\mathrm{Rb}(1)]\sim 480-450$ cm^{-1}) except for methane ($\Delta\nu_m[\mathrm{Rb}(1)]\sim 390$ cm^{-1}) is considered to be due to the effect of collisions localized at a colliding part of the molecular surface, instead of with the whole molecule. Considering that intermolecular repulsion comes from electrostatic interaction between slightly bared nuclei due to distortions of the electronic distribution, $\Delta\nu_m$ may be roughly inversely proportional to the charge density [roughly$\sim(2n+2)/n^{\frac{3}{2}}$] at the contour surface and to the curvature (roughly $\sim 1/n^{\frac{1}{2}}$), where n is the number of carbon atoms in the paraffin molecule. Thus, $\Delta\nu_m\sim n/(2n+2)$, which predicts the smallest $\Delta\nu_m$ for methane. Although this calculation would predict a larger $\Delta\nu_m$ for large n, this effect is presumably suppressed by the larger dispersion energy of the heavier molecule. Actually, when molecules are approaching each other, the valley of potential minimum may tend to orient towards the colliding pair in spite of thermal rotations. This effect would be also one of the reasons for the same value of $\Delta\nu_m$ with butene-2 as with those of normal hydrocarbons.

Although Preston's idea could be used to account for some observed phenomena as described above, one should note that his picture is still far from complete. Several bands associated with an atomic line require several minima and maxima in the potential difference curve, which is hardly to be expected, although the potential curve at close distances must be very complicated. The rapid decrease in $\Delta\nu_m$ of the red and the violet bands for higher member lines and the usual expectation of less steep repulsive potentials for higher excited states may indicate that the potential difference curves tend to have less and less pronounced minima and maxima for higher members, resulting in a monotone curve such as Eq. (52). Then, for highly excited states the potential well may be deep and wide enough for easier molecular formation, which is not observed. Furthermore, the quantum number J may not be good any more, since an angular momentum originated in thermal motion of a perturber, including both translational and, in the case of hydrocarbons, rotational, should be added. The intermolecular field deviates from the spherically symmetric one at close distances when an alkali atom is in the state $^2P_{\frac{1}{2}}$, so that the angular wave function for thermal motion in the ground and excited states are not orthogonal to each other. The angular momentum of thermal motion may be required to change at an electronic transition although this is not obvious without calculation. Red and violet bands are modifications of atomic lines due to thermal motion of perturbers. But why is it possible to broaden the line several times larger in energy units than the energy of the thermal motion per particle? Although generally the statistical theory does not lead to large errors, even when it is not obviously applicable, it has not been considered quantum-mechanically for the case of close encounters. The WKB method[98] loses its validity, when the potential varies rapidly over the wavelength for the motion of a perturber. Also the representation of perturbers as points, instead of wave packets, used in the impact theory becomes worse. The interatomic region responsible for the red and violet bands may be largely affected by quantum-mechanical effects due to the noncommuting properties involved, including the radial distribution function of perturbers at close distances. Thus, this region has been excluded, as a quantum-mechanical region, in the discussion of II.B., C, and D.

D. PRESSURE BROADENING OF MOLECULAR BANDS IN THE INFRARED REGION

The effect of transparent foreign gases in increasing the infrared absorption of a given amount of absorbing gas has been known since 1889.[157a] Beer's law,[157b] however, states that each molecule of a gas absorbs independently and the absorption coefficient is proportional to the concentration of the absorbing gas molecules. This discrepancy has been attributed to molecular collisions in the gas mixture[158] and, later, to collision broadening of individual absorption lines.[159]

A quantitative measurement of the effect of various transparent gases on the intensity of the absorption maximum of N_2O at 4.5 μ and of CO at 4.66 μ ($T=298°$K) was made by Cross and Daniels[160] with instruments of medium resolution. The effective cross sections of foreign gases for increasing infrared absorption thus found are linearly related to the gas kinetic diameters. Likewise the effect was also studied for methane at 7.65 μ, CO_2 at 4.3 μ, and isobutylene at 14.8 μ by Coggeshall and Saier,[161] as shown in Fig. 19. They found that the pressure broadening effects of certain gases on one absorber were characteristic of

[157a] K. Angström, Öfversigt af K. Vetensk. Akad. Förh. 46, 549 (1889); W. W. Watson, J. Phys. Chem. 41, 61 (1937).
[157b] A. v. Beer, Pogg. Ann. 86, 78 (1852).
[158] Eva v. Bahr, Ann. Physik 33, 585 (1910).
[159] C. Füchtbauer, Physik. Z. 12, 722 (1911).
[160] P. C. Cross and F. Daniels, J. Chem. Phys. 2, 6 (1934).
[161] N. D. Coggeshall and E. L. Saier, J. Chem. Phys. 15, 65 (1947).

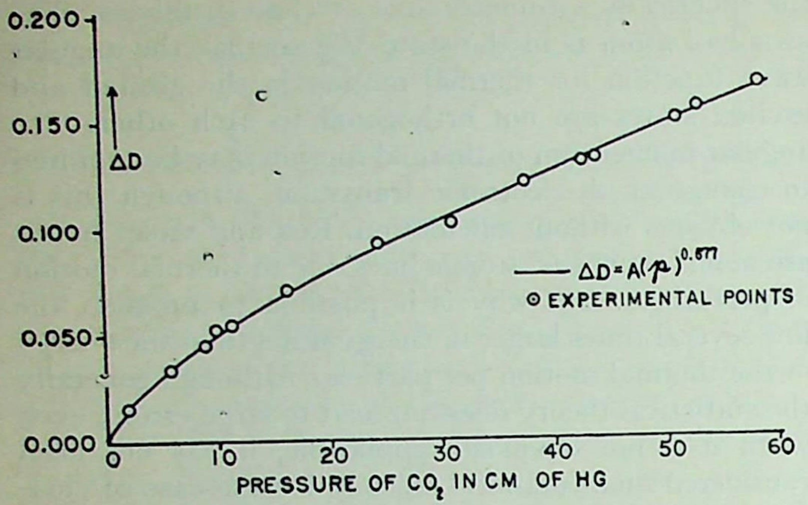


FIG. 19. Increase in optical density of constant pressure of methane as a function of foreign gas pressure (after Coggeshall and Saier[161]).

the nature of the absorber and the wavelength studied, just as the pressure effects on atomic lines.

Usually the absorption coefficient is measured by observing transmittance T_i, when a spectrometer is set to transmit radiation of a certain frequency ν_i. But it actually transmits a frequency band of finite width given by a spectrometer transmission function $f(|\nu-\nu_i|, a)$, where a is the width of the entrance slit. T_i is defined by I_i/I_{0i}, where I_{0i} is the intensity of the radiation transmitted by absorption cell and spectrometer when the cell is empty and I_i is the intensity of the radiation transmitted when the cell is filled with the absorbing gas. If I_ν represents the energy distribution in the radiation reaching the slit through the empty cell,

$$T_i=\frac{I_i}{I_{0i}}=\frac{\int_0^\infty I_\nu \exp(-k(\nu)/l)\, f(|\nu-\nu_i|, a)d\nu}{\int_0^\infty I_\nu f(|\nu-\nu_i|, a)d\nu}, \qquad (142)$$

where l is the absorption length and $k(\nu)$ the absorption coefficient under collision damping, which is given by a general expression summing all single absorption lines m,

$$k(\nu)=\sum_m k_m(\nu)=\sum_m \alpha_m/2\pi^2\tau\left[(\nu-\nu_m)^2+\left(\frac{1}{2\pi\tau}\right)^2\right]. \qquad (143)$$

ν_m is a center frequency of a single line and α_m is the total absorption coefficient of the line m.

For simplicity, if lines are assumed to appear quite isolated and when the frequency interval transmitted is so narrow that I_ν may be considered to be constant over this interval, Eq. (142) leads to

$$T_i=\frac{\int_0^\infty \exp\{-\alpha_m l/2\pi^2\tau[(\nu-\nu_m)^2+(1/2\pi\tau)^2]\} f(|\nu-\nu_i|, a)d\nu}{\int_0^\infty f(|\nu-\nu_i|, a)d\nu}. \qquad (144)$$

For further calculation, a detailed knowledge of the function $f(|\nu-\nu_i|, a)$ is required. Landenberg and Reiche[162] took $f(|\nu-\nu_i|, a)$ to be constant in the frequency interval ν_i-a, ν_i+a and zero outside. Dennison[163] used a Gaussian function of the form, $\exp[-2.77(\nu-\nu_i)^2/a^2]$. Nielsen, Thornton, and Dale[164] assumed a triangular form, with maximum at $\nu=\nu_i$ and extending from ν_i-a to ν_i+a. With one of these assumptions, expansion of the exponentional function in Eq. (144) and integration term by term give, at $\nu=\nu_i$, when $(1/\tau^2)\ll(\alpha l/\tau)<2(\pi a)^2$,

$$T_m=1-2\frac{(\alpha_m l/2\pi\tau)^{\frac{1}{2}}}{a}+\cdots \simeq \exp[-2(\alpha_m l/2\pi\tau)^{\frac{1}{2}}/a], \qquad (145)$$

for a rectangular form and

$$T_m=\exp[-1.879(\alpha_m l/2\pi\tau)^{\frac{1}{2}}/a], \qquad (146)$$

for a Gaussian form. Thus, the result is essentially insensitive to the form of the transmission function. These expressions have been successfully used to account for measurements on individual strong lines. But to get over-all dependence on α_m, τ, l, and a of T_m, Eq. (144) must be numerically integrated.[164]

In infrared bands of polyatomic molecules, the individual lines generally lie so close together that a considerable number of them have their centers in the narrow frequency range transmitted by the spectrometer. The idealized case of an absorption band consisting of an infinite sequence of equidistant identical absorption lines, of which many fall in the frequency band transmitted by the spectrometer, has been treated by W. M. Elsasser.[165] Starting from Eqs. (142) and (143) with $\alpha_m=$ const (α) and $\nu_m=md$ because ν_m distributes with equal spacing d in frequency, he obtained, for small values of $1/d\tau$ (low pressure),

$$T=1-\Phi\left(\frac{(\alpha l/2\tau)^{\frac{1}{2}}}{d}\right). \qquad (147)$$

[162] R. Landenburg and F. Reiche, Ann. Physik 42, 181 (1913).
[163] D. M. Dennison, Phys. Rev. 40, 81 (1932).
[164] Nielsen, Thornton, and Dale, Revs. Modern Phys. 16, 307 (1944).
[165] W. M. Elsasser, Phys. Rev. 54, 126 (1938).

Φ is the probability integral. For large values of $1/d\tau$ (high pressure),

$$T=\exp\left[-\frac{\alpha l}{d}\tanh\frac{1}{d\tau}\right]\cdot J_0\left(i\frac{(\alpha l/d)\tanh(1/d\tau)}{\cosh(1/d\tau)}\right). \quad (148)$$

J_0 is the Bessel function of zeroth order.

The above calculations are only for limiting cases and actual lines are neither quite isolated nor equally spaced in an infinite sequence, but rather are between them. Extensions to more general types of vibration-rotation bands have been given by Matossi and collaborators.[166] However, the above results for idealized cases lead to the correct functional relations between absorption, pressure, and optical density.

The discussion above has been limited to the case of Lorentz broadening. With strongly polar or reactive gases, larger deviation from a simple absorption expressed by Eqs. (147) and (148) may be expected.

When instruments of high resolving power are used, the half-widths of the individual rotational lines can be measured instead of the total band absorption. Herzberg and others[167] investigated the self-broadening of the rotational lines in HCN. The same problem was studied by Kortüm and Verleger[168] for both self-broadening and pressure broadening due to the addition of different foreign gases. The line widths of both the P and the R branch of the 0–3 vibration-rotation band of HCN at $1.0380\,\mu$ were measured. The self-broadening of the HCN band over the pressure range 75 to 550 mm Hg is shown in Fig. 20. It is to be noted that the half-widths for both branches reach a maximum at the density maximum of the band ($J\simeq 8$), and decrease with both decreasing and increasing J's.

The broadened lines of linear polar molecules are somewhat less well understood because of the difficult calculation of the interaction energy involved. One

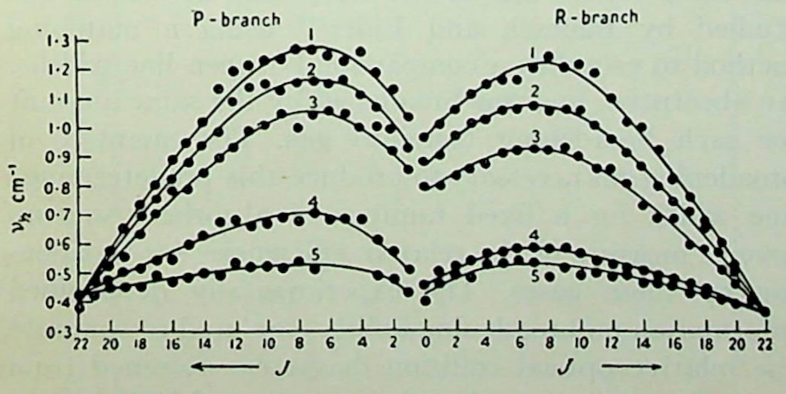


FIG. 20. The half-width values of the absorption lines of the 0–3 band of HCN in relation to the rotational quantum number J of the P and R branch for different pressures of the absorption gas (after Kortüm and Verleger[168]). 1: $p=550$ mm; 2: $p=450$ mm; 3: $p=300$ mm; 4: $p=150$ mm; 5: $p=75$ mm.

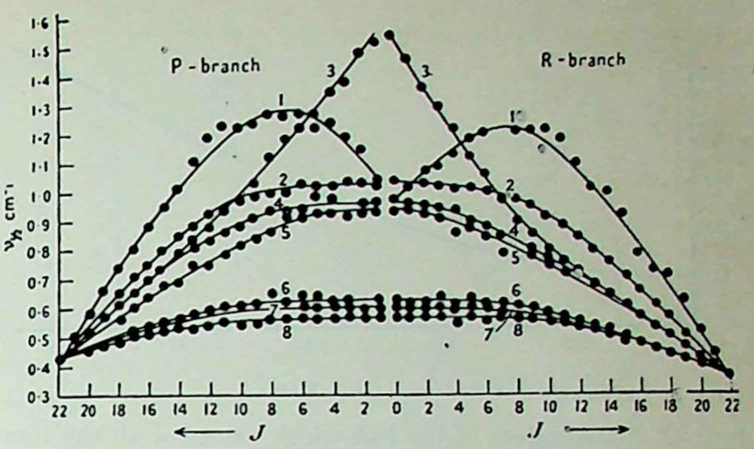


FIG. 21. The half-width values of the absorption lines of the 0–3 band of HCN in relation to the J's of P and R branches when various foreign gases are added (after Kortüm and Verleger[168]).

1: HCN/HCN	5: $HCN/(CH_3)_2O$
2: HCN/DCN	6: HCN/H_2S
3: HCN/C_2H_5Cl	7: HCN/N_2O
4: HCN/SO_2	8: HCN/N_2

should expect at least two types of interactions between HCN molecules, rotational and dipole-dipole, although the rotating dipoles should not have any permanent dipole moment on the average. As described in Sec. II. A. 2.b.i, the dipole-dipole interaction for lower rotational levels must obey the R^{-3} law, since the interaction between molecules due to the thermal motion is easily larger than the level spacing $\hbar^2 J/I$ for such levels, I being the moment of inertia. Thus, a permanent dipole moment may appear. Under these considerations, the interaction energy difference in the transition $J-1\rightarrow J$, will be $0.2\ \mu\mu'/JR^3$, as given in Eq. (17), for lower rotational levels. Hence, through Eq. (55), a larger half-width is expected for lines with smaller J values. In addition, for self-pressure, the two interacting molecules, with $\Delta J=\pm 1$ either before or after absorption, undergo resonance, according to Eq. (13), either before or after absorption respectively. This resonance causes another half-width which depends on J through Eq. (13) and on its statistical population, giving a maximum around $J=8$. The total half-width will be given approximately by combining the effects due to rotational resonance and dipole-dipole interaction.

The rate of decrease in half-width for increasing J is greater for higher pressures, and the absolute value of half-width for the R branch is smaller than those for the P branch, as theoretically predicted by Lindholm[167] through London's formula.[36] The average half-width values for all band lines were strictly proportional to the pressure for the pressure range studied. The slope was about $9.1\cdot 10^{-4}$ cm^{-1}/mm Hg.

Figure 21 gives the half-width values of the absorption lines of the 0–3 band of HCN in relation to J of the P and R branches when various foreign gases were added. The HCN partial pressure was kept at 75 mm and the total pressure at 550 mm. It is to be noted that, under the pressure of foreign gases, the half-widths were not greatest at the intensity maximum of

[166] Matossi, Mayer, and Raucher, Phys. Rev. **76**, 760 (1949).
[167] G. Herzberg and J. W. T. Spinks, Proc. Roy. Soc. (London) **A147**, 434 (1934); S. D. Cornell, Phys. Rev. **51**, 739 (1937); E. Lindholm, Z. Physik **109**, 223 (1938); **113**, 596 (1939).
[168] G. Kortüm and H. Verleger, Proc. Phys. Soc. (London) **63**, 462 (1950).

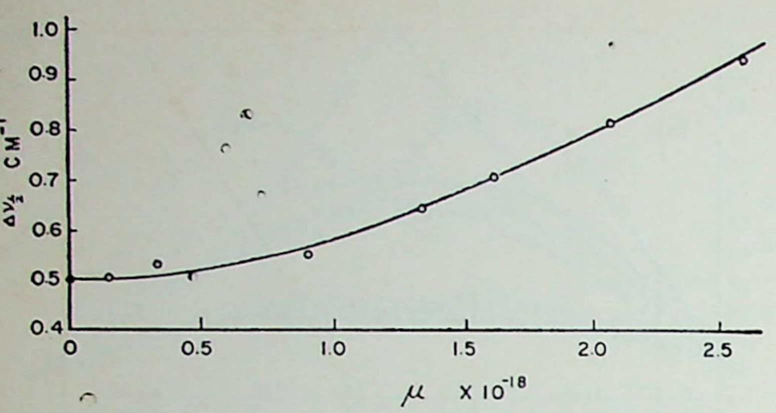


FIG. 22. Mean values of the half-width values of all the rotational lines of the HCN 0–3 band in relation to the dipole moment of the added gas (after Kortüm and Verleger[168]).

the band, but reached a maximum for lines which resulted from transitions between energy states of low J values. This is to be expected from Eq. (17) and the discussion in the previous paragraph, since no resonance energy is assigned in this case. The effect that the broadening decreases in both branches for increasing J was particularly apparent for ethyl chloride. The effect that the half-width due to DCN (with the largest dipole moment of all) is smaller than that due to ethyl chloride was considered possibly due to the fact that the hydrogen bonds had a strong affinity for chain association. In particular, the resonance effect predominant for the HCN case should be reduced for DCN, because of the different magnitudes of moments of inertia. Figure 22 is a plot of the average half-width of all the lines of a band *vs* $\mu\mu'$, where μ is the dipole moments of the added gases and μ' that of HCN. It follows from the curve that the half-widths first changed slowly for small μ and then the variation was amost linear with $\mu\mu'$ when $\mu > 0.75$.

The rotational lines of the near infrared band of HCl, H_2O and CO_2 were broadened to about the same extent and the line width could be accounted for by Lorentz theory, *viz.*, approximate proportionality between the width and the pressure and with intensity distributions of the dispersion form.

Coulon[169,170] examined the perturbation of the infrared fundamental bands of HCl (3000 cm^{-1}) by N_2, A, He, H_2 and O_2 and of the 2100 cm^{-1} band of CO by N_2 and A compressed up to 1000 atmospheres, using a Perkin-Elmer spectrometer with LiF prism. With He, H_2, or argon the rotational structure was still visible for the P branch of HCl at r.d. 350. The two maxima for the P and the R branches were obvious even when the argon r.d. was 450, the highest attained. With nitrogen the rotational structure disappeared around r.d. 170. The perturbation grew stronger in the order He, H_2, A, O_2, N_2. An absorption maximum very close to the forbidden Q branch of HCl was observed as induced by the nitrogen pressure. The intensity of this Q maximum increased rapidly with pressure. As is characteristic of collision-induced absorption,[171,172] the total intensity of the band, $I = \int \alpha(\nu)d\nu$, increased linearly with the density of the gases beyond 50 r.d. (but more rapidly with N_2 than with a spherical atom A). For CO (with a smaller dipole moment than that of HCl) the changes in intensity are smaller than for HCl.

In the case of the vibration-rotation bands of some other polystomic molecules such as NH_3, no J dependence of this line broadening could be established, although there was a definite increase in line width with pressure. Other molecules such as CH_4 gave sharp lines under similar conditions.[173]

When the rotational half-width of molecules were very small, so small that the direct measurement of the line shape was beyond the resolving power attainable with most spectrographic instruments, it could be estimated indirectly by studying the dependence of apparent absorption on total pressure.[174] From Eq. (147) $T \sim 1 - 2(\alpha l/2T)^{\frac{1}{2}}/d$ or $-dT/d(p)^{\frac{1}{2}} = 2(\alpha l/2\tau_0)^{\frac{1}{2}}/d$ where $1/\tau = (1/\tau_0)p$, p being a total pressure and $1/\tau_0$ a number of optical collision per unit time at a unit pressure. Thus, the last expression, i.e. the slope of a T *vs* $p^{\frac{1}{2}}$ plot can be used to determine half-width $1/2\pi\tau_0$ at unit pressure. Thus, Penner and Weber showed that the rotational half-width of the fundamental line of CO in the presence of H_2 or A was 0.077 cm^{-1} per atmosphere or 0.040 cm^{-1} per atmosphere, respectively. Similarly for H_2 the first overtone of CO showed a half-width of 0.063 cm^{-1} per atmosphere. These results are somewhat smaller than the values 0.10 cm^{-1} atmos^{-1} and 0.12 cm^{-1} atmos^{-1} for the fundamental and first overtone line with air.[175]

The pressure broadening of individual pure rotational lines, *viz.*, the $J = 2 \to 3$ line of the HCl band at 3.4 μ and the $J = 4 \to 5$ line of the CH_4 band at 3.25 μ, was studied by Benesch and Elder[176] using a matching method to establish a comparison between line widths. An absorption line was broadened by the same amount for each broadening (foreign) gas. The quantity of broadening gas necessary to produce this predetermined line width for a fixed number of absorbers was an inverse measure of the relative efficiencies for broadening by these gases. The experimentally determined information on broadening ability was used to compute the relative optical collision diameters obtained from pressure broadening of an infrared line of HCl against the values of absolute collision diameters obtained from

[169] Coulon, Oksengorn, and Robin, Compt. rend. **236**, 1481 (1953).
[170] Coulon, Galatry, Oskengorn, Robin, and Vodar, J. phys. radium **15**, 58 (1954).
[171] Crawford, Welsch, and Locke, Phys. Rev. **75**, 1607 (1949); M. Mizushima, Phys. Rev. **76**, 1268 (1949); **77**, 149 (1950).
[172] Kiyama, Minomura, and Ozawa, Proc. Japan Acad. **30**, 758 (1954).
[173] H. Verleger, Physik. Z. **38**, 83 (1937).
[174] For instance, A. M. Thorndike, J. Chem. Phys. **16**, 211 (1948); S. S. Penner and D. Weber, *ibid.* **19**, 1351 (1951).
[175] L. A. Matheson, Phys. Rev. **40**, 813 (1932).
[176] W. Benesch and T. Elder, Phys. Rev. **91**, 308 (1953).

pressure broadening of a microwave absorption line of NH_3 (by Smith and Howard[42] and Hill and Smith[177]) using the same foreign gases, as shown in Fig. 23. The correlation indicated in the figure emphasizes the fundamental similarity of pressure broadening results in the two distinct spectral regions. In this particular case, not only the frequency (3000:1) was enormously different (leading to a corresponding disparity between the duration of collision and the period of the absorbed radiation) but the dipole moments were different, and, furthermore, one line (HCl line) originated in molecular vibration-rotation and the other (NH_3 line) resulted from the inversion of the ammonia molecule.

To show the effect of a dipole moment in the absorbing molecule Benesch and Elder[176] also used nonpolar CH_4 as absorber. The relative collision diameters obtained for HCl absorption were plotted against those for CH_4 absorption as shown in Fig. 24. The points fall nearly along two straight lines, except for HCl which displayed a higher collision cross section due to resonance interaction. The lower lines belong to points for molecules with a high degree of symmetry, *viz.*, the five rare gases and SF_6. This line, like that of Fig. 23, nearly passes through the origin, showing that the interaction is probably dependent on the polarizability of the perturbing molecules. The strength of interaction was not constantly enhanced by the dipole moment in the absorbing molecule.

The much greater slope of the upper line indicates that the ratio of the interaction strength between the same molecules and CH_4 is much greater, due to the influence of the dipole moment of HCl. So additional

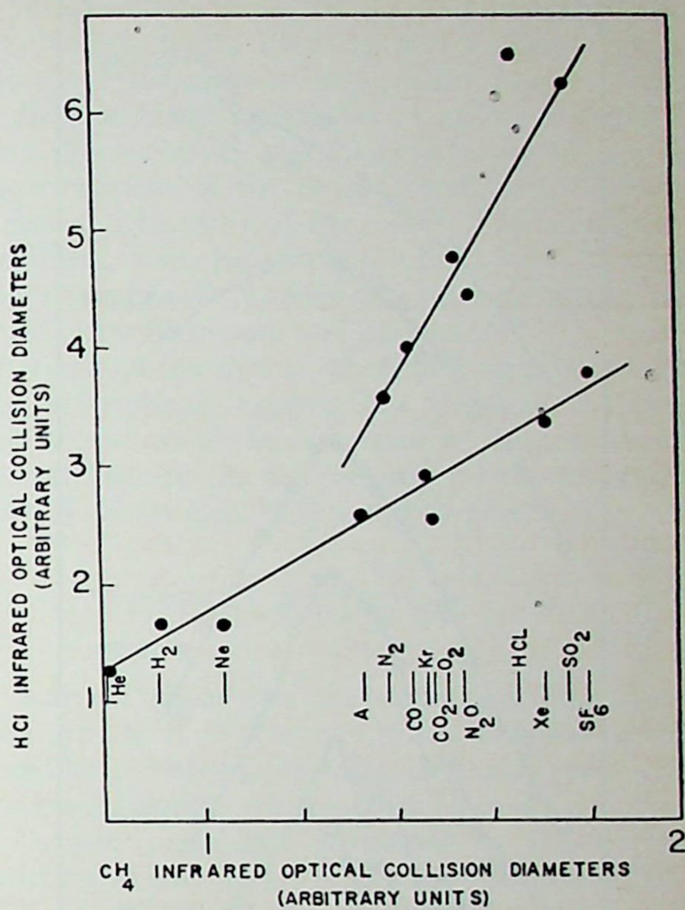


FIG. 24. Relative optical collision diameters obtained by pressure broadening of an HCl infrared absorption line compared with those obtained by pressure broadening of a CH_4 infrared absorption line with the same gases (after Benesch and Elder[176]).

interaction strength due to this dipole moment was added to the common value shown in the lower line. This investigation pointed out the source of difficulty inherent in previous interpretations of experimental data. While a single type of long range interaction between dipole moment, induced dipole moment and quadrupole moment of the absorber and the perturber may be capable of explaining the behavior of one particular group of molecules, at least two different types of interactions were involved in the pressure broadening of a dipole absorber when a variety of perturbers was used.

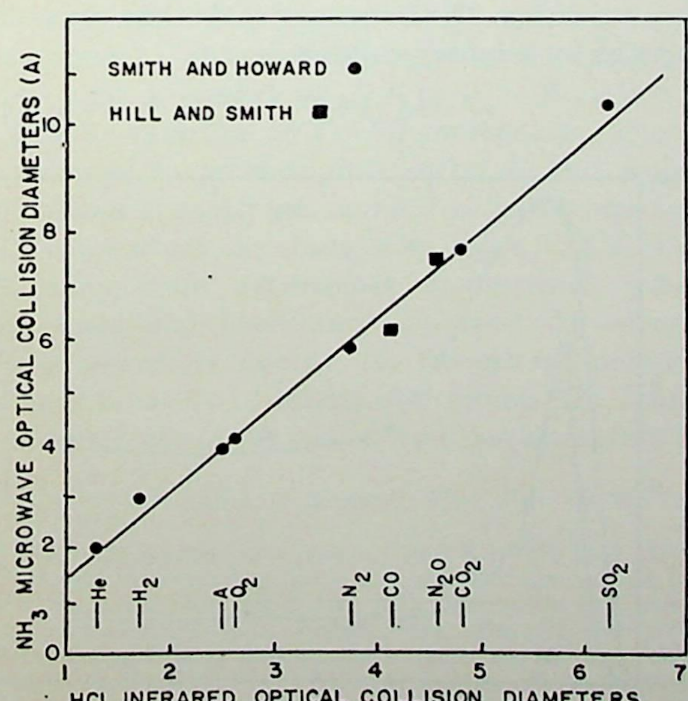


FIG. 23. Optical collision diameters obtained from broadening of a microwave absorption line of NH_3 compared with relative optical collision diameters obtained from broadening of an infrared line of HCl (after Benesch and Elder[176]).

E. PRESSURE BROADENING OF LINES IN THE MICROWAVE REGION

More accurate measurements of the absorption intensities can be made in the centimeter wave region than in the optical or the infrared regions. Thus, the shape of the spectral lines and the variation of their widths have been studied over a range of pressures from atmospheric down to 0.01 mm Hg or less. For low pressures, effects due to multiple encounters can be neglected and the Van Vleck-Weisskopf expression (36a) will be applicable, but the theoretical analysis is not yet refined well enough to give a complete interpretation of experimental data.

Considerable work has been done on self-broaden-

[177] R. M. Hill and W. V. Smith, Phys. Rev. 82, 451 (1951).

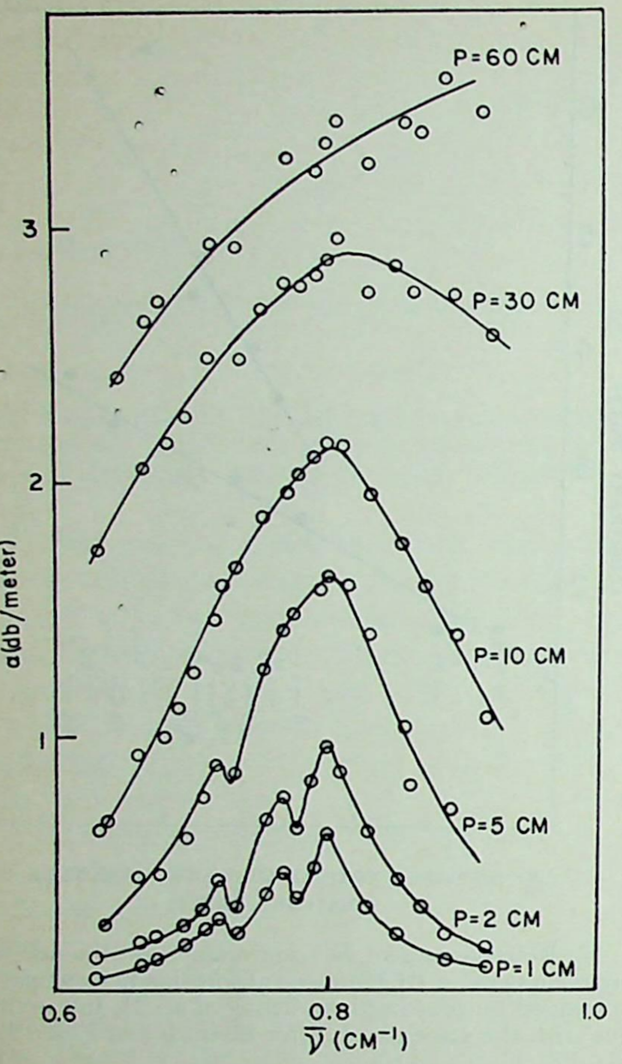


FIG. 25. Absorption curves for ammonia at various pressures (after Bleaney and Penrose[179]).

ing[178] such as the broadening of the inversion spectrum of ammonia[42,177,179] ($\rho \sim 14$ A for (3.3) line), oxygen[180] ($\rho \sim 5$ A for $J=6 \rightarrow 8$ line), water[181] ($\rho \sim 10$ A for $J=5_{-1} \rightarrow 6_{-5}$ line), methyl halides[182] ($\rho \sim 8$ A) and of rather simple linear molecules[183] ($\rho \sim 10$–20 A). The values of ρ, the collision diameter, is indicated in each case. Figure 25 is a sample curve for ammonia at various pressures. The lines become sufficiently broad to obscure the fine structure at a pressure of 10 cm Hg.

At low pressures, say, between 0.5 mm and 10 cm Hg, all results obtained by various investigators confirm that the line widths vary linearly with pressure, and there is no observable shift. For ammonia, however, when the pressure is high (above 30 cm Hg) the experimental curve[179] shows a definite shift to lower frequencies as shown in Fig. 26, and the resonance frequency becomes substantially zero for pressures above two atmospheres. The line width no longer increases linearly with pressure but rather decreases with pressure, but above four atmospheres, again, it increases linearly with pressure.

This pressure variation for ammonia, particularly of the shift, has been explained in a semiquantitative way.[184] Margenau,[35] applying the variational method, calculated the shift of and the transition probabilities between energy levels of two and three NH_3 molecules under dipole-dipole interaction in a simple model. He showed that the inversion line split, due to the perturbation, into respectively two or three lines. In either case, only one line appears with decreasing frequency but with increasing intensity with perturbation at the expense of the rest. Thus, the shift is obvious and the presence of two or more frequencies with changing intensities will destroy the proportionality of line widths with pressure and narrowing may take place, although no rigorous theoretical treatment about this effect has appeared. If the pressure becomes high, the most intense line component will accentuate its frequency reduction. The frequency of this line component changes as for a binary collision,

$$\nu_0' = \nu_0[(1+\lambda^2)^{\frac{1}{2}} - \lambda], \tag{149}$$

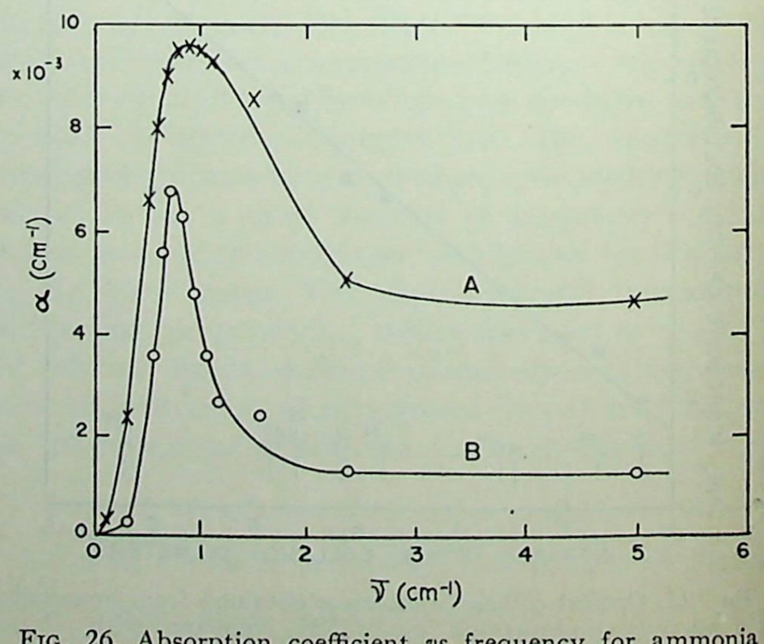


FIG. 26. Absorption coefficient *vs* frequency for ammonia *A*: 90 cm Hg; *B*: 30 cm Hg (after Bleaney and Loubser[179]).

[178] Gordy, Smith, and Trambarulo, *Microwave Spectroscopy* (John Wiley and Sons, Inc., New York, 1953), p. 185; C. H. Townes and A. L. Schawlow, *Microwave Spectroscopy* (McGraw-Hill Book Company, Inc., New York, 1955).

[179] C. H. Townes, Phys. Rev. **70**, 665 (1946); B. Bleaney and R. P. Penrose, Proc. Phys. Soc. (London) **59**, 418 (1947); Proc. Roy. Soc. (London) **A189**, 358 (1947); B. Bleaney and J. H. N. Loubser, Nature **161**, 522 (1948); Proc. Phys. Soc. (London) **A63**, 483 (1950); D. F. Smith, Phys. Rev. **74**, 506 (1948); I. R. Weingarten, thesis, Columbia University (1948), and reference 47.

[180] Anderson, Smith, and Gordy, Phys. Rev. **82**, 264 (1951); R. Beringer, Phys. Rev. **70**, 53 (1946); J. H. Van Vleck, Phys. Rev. **71**, 413 (1947); Strandberg, Meng, and Ingersoll, Phys. Rev. **75**, 1524 (1949); H. R. L. Lamont, Phys. Rev. **74**, 353 (1948); J. O. Artman and J. P. Gordon, Phys. Rev. **87**, 277 (1952), and reference 191.

[181] (a) G. E. Becker and S. H. Autler, Phys. Rev. **70**, 300 (1946); (b) C. H. Townes and F. R. Merritt, Phys. Rev. **70**, 558 (1946).

[182] B. Bleaney and J. H. N. Loubser, reference 179; Gillam, Edwards, and Gordy, Phys. Rev. **75**, 1014 (1949).

[183] ICI, Townes, Merritt, and Wright, Phys. Rev. **73**, 1334 (1948); OCS, Townes, Holden, and Merritt, Phys. Rev. **74**, 1113 (1948); C. M. Johnson and D. M. Slager, Phys. Rev. **87**, 677 (1952); and references 47 and 192; ICN, ClCN, BrCN, Smith, Gordy, Simmons, and Smith, Phys. Rev. **75**, 260 (1949); and reference 193; CH_2O, R. B. Lawrence and M. W. P. Strandberg, Phys. Rev. **83**, 363 (1951).

[184] P. W. Anderson, Phys. Rev. **75**, 1450 (1949); H. Margenau, Phys. Rev. **76**, 121 (1949).

where $\lambda \sim 2\mu^2 N/h\nu_0 c$, ν_0 being the unperturbed frequency of the inversion line. μ is the permanent dipole moment, and N the number of molecules per cc. For $\lambda \gg 1$, ν_0' equals substantially zero and the absorption should change from the resonant absorption to that of the Debye type.

The experimental confirmation of Eq. (149) may be attributed to the work by Birnbaum and Maryott.[185] The completely deuterated ammonia ND_3 is expected to behave in the same manner as NH_3 under pressure. However, since the resonance frequency ν_0 of the ND_3 inversion line is near 1600 Mc/sec[186] and 14.8 times less than that of NH_3, the shift in the inversion frequency should occur at pressures about 15-fold lower than in the case of NH_3 according to Eq. (149). They observed this relationship with a good agreement using the results for NH_3 by Bleaney and Loubser.[179] At higher pressures (above 6 cm Hg) the resonance frequency of ND_3 drops to zero more rapidly than predicted by Eq. (149).

Measurement of the half-widths of 14 resolved magnetic absorption[187] lines in the 5-mm wavelength region for oxygen were made by Anderson, Smith and Gordy[188] and others.[189]. Line-breadth parameters††† obtained ranged from 0.0319 to 0.0516 cm^{-1}/atmos with a population weighted mean of 0.0475 cm^{-1}/atmos. On the other hand, the half-width for the rotational transition at 2.5 mm wavelength was found by Hill and Gordy[190] as $\Delta\nu_{\frac{1}{2}}' = 0.064$ cm^{-1}/atmos at $T = 300$°K and 0.09 cm^{-1}/atmos at 195°K.

The difference between these half-width parameters can be understood as caused by the temperature difference. In this case, $\Delta\nu_{\frac{1}{2}}$, in units of cm^{-1} atmos^{-1}, will vary approximately as T^{-1} according to their measurement, although Beringer and Castle[191] observed that it varied as $T^{-\frac{3}{4}}$ to T^{-1}. Equation (56) predicts its variation as $T^{-\frac{3}{2}}$ for constant $\Delta\gamma$, since N is proportional to pressure divided by T. But since the populations of rotational levels J of perturbers change with temperature, $\Delta\gamma$ changes its value with T. Hence, this temperature dependence is extremely complex when rotational levels are involved. However, its observation offers one of the important methods of finding the nature of interaction forces. The temperature dependence of the optical collision diameters, $T^{-0.2}$ for OCS[192] and $T^{-0.5}$ for BrCN,[193] has been theoretically analyzed by Smith, Lackner, and Volker.[43] They find that the T dependence requires the introduction of a quadrupole-dipole interaction in addition to the first-order dipole-dipole, and have determined the quadrupole moment of BrCN as $Q/e = 5.7 \times 10^{-16}$ cm^2. For ammonia, Bleaney and Penrose[179,194] obtained the T^{-1} law. Then, with Eq. (35), the absorption at wings is proportional to T^{-3}, which has been confirmed experimentally by Birnbaum and Maryott.[195]

Because of the special origin of a spectral line (magnetic absorption), oxygen, like ammonia[194] (inversion spectra), is one of the rare cases which has microwave spectral lines in the significantly populated rotational states K. Since Eqs. (23) and (56) give[37]

$$\Delta\nu_{\frac{1}{2}} = 1.33 \times 10^{15} Q (1/K \pm \tfrac{3}{8}K^2)^{\frac{1}{2}} \text{ cm}^{-1} \quad \text{for} \quad \begin{matrix} K+1 \rightarrow K \\ K-1 \rightarrow K \end{matrix}$$

for high K values and the dispersion force gives the same type of K dependence,[64] the variation of the observed line width as a function of K may reveal again a rotational resonance interaction, with its maximum at the largest population predicted by the Boltzmann statistics, as mentioned before for the infrared lines of HCN (Fig. 20). This indication of resonance interation might be checked by changing the temperature of the gas, since the change in K value for maximum population thus resulting could cause the shift of the K value corresponding to the maximum half-width.

Pressure broadening due to foreign gases in the microwave region has been studied experimentally with ammonia, water and oxygen. Bleaney and Penrose[194] and, later, Smith and Howard[42] and Hill and Smith[177] measured the collision cross sections of the $(J;K) = (3,3)$ line at 0.796 cm^{-1} of the inversion spectrum of ammonia for binary mixtures of seventeen different nonpolar and polar gases. The results are tabulated in Table VI, showing larger collision cross sections for polar gases than otherwise. The ammonia partial pressure was between 0.2 to 3 mm Hg (0.02 to 0.04 mm Hg for Howard and Smith[42]) and the concentration of foreign gas to ammonia was about 10:1. The collision cross section was computed from the measurement of the absorption coefficient α at the center of the line, which is related to half-width $\Delta\nu_{\frac{1}{2}}$ by Eq. (36a), neglecting effects[178] of saturation[50,51] given by Eq. (36b) and collisions with the walls. With $\nu = \nu_0$, Eq. (36a) leads to

$$\alpha = \frac{8\pi^2 |\mu_{JK}|^2 \nu_0^2 N_{JK}}{3kT\Delta\nu_{\frac{1}{2}}} \text{ cm}^{-1}, \tag{150}$$

[185] G. Birnbaum and A. A. Maryott, Phys. Rev. **89**, 895 (1953); **92**, 270 (1953).

[186] Nuckolls, Rueger, and Lyons, Phys. Rev. **89**, 1101 (1953).

[187] Oxygen molecule has no electric dipole moment, but it does have a magnetic moment of two Bohr magneton and gives rise to magnetic absorption. R. Schlapp, Phys. Rev. **51**, 342 (1937).

[188] Anderson, Smith, and Gordy, Phys. Rev. **87**, 561 (1952).

[189] B. V. Gokhale and M. W. P. Strandberg, Phys. Rev. **84**, 844 (1951).

††† Line-breadth parameter is defined as the half-width measured at half-intensity (the half-intensity *half-width*), and will be indicated as $\Delta\nu'_{\frac{1}{2}}$. The half-width, $\Delta\nu_{\frac{1}{2}}$ (without a prime) as used in Sec. III(1) is half-intensity *width*, and is, therefore, twice larger.

[190] R. M. Hill and W. Gordy, (unpublished).

[191] R. Beringer and J. G. Castle, Jr., Phys. Rev. **81**, 82 (1951).

[192] Feeney, Lackner, Moser, and Smith, J. Chem. Phys. **22**, 79 (1954).

[193] Trambarulo, Lackner, Moser, and Feeny, Phys. Rev. **95**, 622(A) (1954).

[194] B. Bleaney and R. P. Penrose, Proc. Phys. Soc. (London) **60**, 540 (1948).

[195] G. Birnbaum and A. A. Maryott, J. Chem. Phys. **21**, 1774 (1953).

Table VI. Optical collision diameters for mixtures of various gases in the microwave region.

Absorber and line	Perturber	$\mu(\bar{\mu})$[a] Debye	Molecular weight of perturber	Optical collision diameter A Exp.	Optical collision diameter A Theor.[b]	Kinetic collision diameters[c]	Reference
NH_3 (3,3)	HCN(2.59)[d]	2.96(0)	27.0	9.95			g
	ClCN(−)	2.80(0)	61.5	11.9			
	CH_3Cl(4.56)	1.87(0.47)	50.5	11.3			
	SO_2(3.72)	1.6(0)	64.1	10.4			
	CH_2Cl_2(6.48)	1.59(0)	85.0	10.3			
	NH_3(2.26)	1.44(0.78)	17.0	13.8[e]	14.0[f]	4.43	h
	$CHCl_3$(8.23)	0.95(0.57)	119.0	13.7			g
	OCS(−)	0.71(0)	60.1	7.54			
	N_2O(3.00)	0.25(0)	44.0	7.32		4.35	i
	CS_2(8.74)	0(0)	76.1	7.5–7.72			g, h
	CO_2(2.65)	0(0)	44.0	7.59		4.51	g
	CCl_4(10.5)	0(0)	154.0	7.20			
	O_2(1.60)	0(0)	32.0	4.85–3.86	3.35	4.02	g, h
	N_2(1.76)	0(0)	28.0	6.4–5.54	3.39	3.4	
	H_2(0.79)	0(0)	2.0	3.50–2.95	2.57	3.58	
	A(1.63)	0(0)	39.9	4.6–3.73	3.42	4.04	
	He(0.22)	0(0)	4.0	2.35–2.00	2.15	3.20	
H_2O ($J=5_{-1}\to 6_{-5}$)	Air	0(0)		5.4		3.5	j
	H_2O(1.5)	1.94	18.0	10.0		10.2	
O_2 ($J=8\to 9$)	NH_3(2.26)	1.44(0.78)	17.0	4.3		4.02	k
	N_2O(3.00)	0.25(0)	44.0	4.4		4.15	
	O_2(1.60)	0(0)	32.0	3.5–5		3.61	
	N_2(1.76)	0(0)	28.0	2.8–4.5		3.68	
	A(1.63)	0(0)	39.9	3.6		3.63	

[a] μ: Static dipole moment of a perturber. $\bar{\mu}$: Weighted average of dipole moment over population of rotational levels.
[b] See reference 41.
[c] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938).
[d] Polarizability (in $cm^3 \times 10^{-24}$) averaged over three directions See Landolt-Börnstein (Springer-Verlag, 1951), I. Band, 3. Teil, p. 511.
[e] For low pressure (7.7 A for pressures more than 4 atmos).
[f] See reference 52.
[g] See reference 176.
[h] See reference 194.
[i] See reference 177.
[j] See reference 181b.
[k] See reference 188.

for relatively sharp lines as are commonly observed in the microwave region. Here,

$|\mu_{JK}|^2$ = square of dipole moment associated with the line (J,K).

ν_0 = wave number of the line in cm^{-1}.

N_{JK} = number of ammonia molecules per cc in rotational level (JK).

$\Delta\nu_{\frac{1}{2}}$ = the half-width in $cm^{-1} = f_{12}/\pi c$, f being the collision frequency between the foreign gas and an ammonia molecule in state (3,3); c, velocity of light.

$$f_{12}=2N_2S_{12}\left(\frac{2RT}{\pi}\right)^{\frac{1}{2}}\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^{\frac{1}{2}},$$

N_2 being the number of foreign gas molecules, M_1 and M_2 the molecular weights, and S_{12} the cross sections for ammonia-foreign gas collisions, which is equal to $\pi\rho_{12}^2$ where ρ_{12} is the collision diameter. Since S_{12} is proportional to f_{12}, which is in turn proportional to $\Delta\nu_{\frac{1}{2}}$,‡‡‡ S_{12} is also a measure of the effect of collision broadening of the (3,3) line produced by various gases. If the order of efficiencies for broadening by these foreign gases is that shown in Fig. 23, the result is in perfect agreement with that obtained with infrared spectra. No shifts of absorption line frequency were observed. These results show that the impact theory is a good approximation under low pressures, as seen in Table VI. However, it might be noticed here that although the idea of the collision diameters is very often useful, for instance, for the determination of the force law, the force laws in many cases are so complicated that the collision diameters are hardly meaningful without detailed discussion.

Becker and Autler[181] made accurate measurements on the broadening of the water vapor line at 0.742 cm^{-1} for mixtures of water vapor and air at 45°C. The half-width $\Delta\nu_{\frac{1}{2}}'=0.087\pm0.001$ cm^{-1} for zero concentration of water in air and, $\Delta\nu_{\frac{1}{2}}'=0.107\pm0.001$ cm^{-1} at a concentration of 50 g/m^3 of water. The values of collision diameter are shown in Table VI. The value of the half-width is much smaller than for NH_3 because a water

‡‡‡ The relation between $\Delta\nu_{\frac{1}{2}}$ and collision diameters is $\Delta\nu_{\frac{1}{2}}=\sqrt{2}n_1v_1\rho_1^2+n_2\bar{v}_{12}\rho_{12}^2$ where n_1 is the molecular density (no. of absorbers per cc) $\bar{v}_{12}=(v_1^2+v_2^2)^{\frac{1}{2}}$, the root-mean-square velocity of the perturbing molecule with respect to the ammonia molecules, and ρ_{12} the collision diameter of the perturbing gas.

molecule is an asymmetrical top. Its dipole moment will have a fairly small steady component in the two rotational levels between which the transition takes place.

Beringer and others[180] measured the integrated high pressure absorption band of oxygen at 2 cm^{-1} mixed with nitrogen and showed that the half-width at atmospheric pressure should be around 0.04 cm^{-1}, giving a collision diameter 3.8 A.

The effect of various concentrations of foreign gases on the width of the oxygen absorption lines was studied by Anderson, Smith and Gordy[188] by measuring the half-widths of the rotational line ($J=8\rightarrow 9$) at a fixed oxygen partial pressure (1.5 mm Hg) and variable total pressure of the mixtures of gases, as shown in Fig. 27. The chosen foreign gases NH_3, N_2O, N_2, and A, represent, respectively, large dipole moment, small dipole moment, large quadrupole moment and electrical inactivity. Since the half-width in Mc of O_2 with O_2 is larger than that with N_2, whose quadrupole moment is expected to be three times larger than that of O_2, the nonresonant quadrupole-quadrupole moment seems to be relatively ineffective in broadening spectral lines.

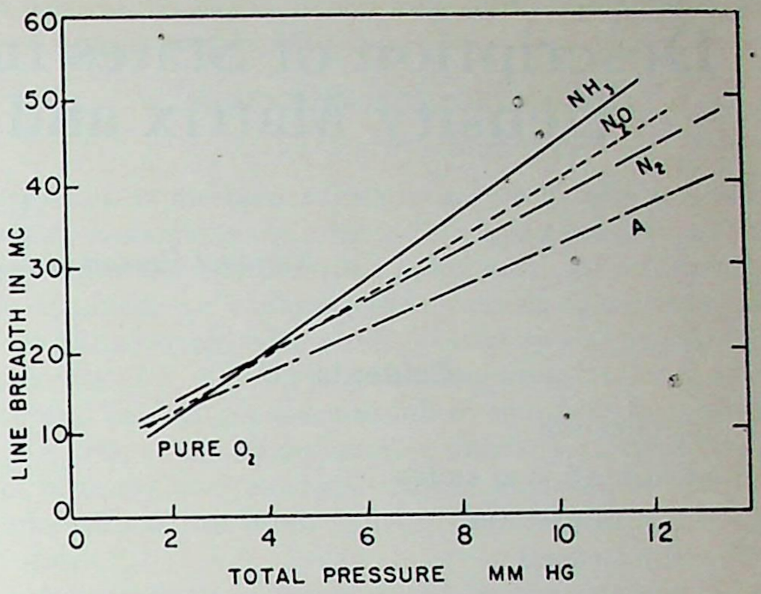


FIG. 27. The broadening of an oxygen line by various foreign gases. $J=8\rightarrow 9$. The partial pressure of O_2 is 1.5 mm Hg (after Anderson, Smith, and Gordy[188]).

In general, since the strong long-range forces give an optical collision diameter larger than a kinetic collision diameter with momentum transfer for the corresponding system, we may conclude that collisions take place at most without deflections of flying path. The theory seems to be satisfactory for the system under interactions of such forces. The relations have been used to obtain force constants or quadrupole moments[42,66] as stated in Sec. II. A.2.d.

The authors take great pleasure in acknowledging the encouragement and very valuable criticism of Dr. Henry Margenau, Professor at Yale University, and Dr. R. T. Ellickson, Professor at the University of Oregon, during the preparation of the manuscript.

Description of States in Quantum Mechanics by Density Matrix and Operator Techniques

U. Fano

National Bureau of Standards, Washington, D. C.

Contents

1. INTRODUCTION

To describe a phenomenon correctly and conveniently, one wants to represent the states of physical systems by parameters that are logically consistent and also have a familiar, operational significance. Difficulties in achieving this goal have come up, for example, in the study of atomic phenomena involving polarization, spin orientations, and angular correlations. In these phenomena, and in many others, the experimental procedure does not usually analyze all relevant variables to the maximum extent consistent with quantum mechanics. We refer to states of "less than maximum information," to indicate for example that we know less about the spin orientation of the protons in a drop of water than we would if each proton were in a state characterized by a magnetic quantum number.

The difficulty in familiarizing oneself with states of less-than-maximum information leads, for example, to ask: "Given a beam of unpolarized electrons, should one think of each electron as having a definite spin orientation?" Questions of this kind are semantically slippery, of course, and should be argued in terms of some specific experimental arrangement. Even so, their clarification requires some effort. This writer regards the picture of a mixture of definite orientations as unrealistic because the choice of alternate orientations is not unique. This is shown, for example, by considering the analogous problem of γ-ray polarization in the phenomenon of two-photon positron annihilation. No polarization of the γ rays is observed with a single detector A, sensitive to polarization but which receives only one photon from any pair. Now, operate A in coincidence with another detector B sensitive to linear polarization and actuated by the other photon of a pair; A will detect a partial linear polarization. If, on the other hand, B is sensitive to circular polarization, A detects a partial circular polarization. Having found that the polarization observed at A depends qualitatively on observations performed at B, one has little basis for classifying the "polarization of individual photons" detected by A, as either linear or circular.

The difficulties met in these discussions derive in part from the circumstance that quantum mechanics has been mostly concerned with the "pure" states of systems, such as states of definite spin orientation, which are represented by state vectors ψ. States with less than maximum information, represented by density matrices ρ, have been considered primarily in statistical me-

chanics and their discussion has been influenced by the historical background in this field. Moreover, most atomic physicists have not had much opportunity to become familiar with the representation of states by density matrices.[1]

Calculations can be, and usually have been, performed without reference to density matrices even when maximum information is not available. This is done by calculating as though a maximum of variables were observed and then summing (or averaging) over the eigenstates of the unobserved variables. On the other hand, when the states are represented by density matrices one can often expedite the calculations and avoid the introduction of unnecessary variables, which is of particular value for the treatment of many-body problems.

The variables to be dispensed with include the arbitrary phase of the state vector ψ which represents the state of the whole system.

It is also important that the density matrix of a system can be easily expressed in terms of the mean values of observables. Identifying a state by means of such physical parameters brings out the operational basis of the theory and helps in forming a mental picture.

In the course of a phenomenon, the variations of density matrices, or of equivalent sets of parameters, can be represented and analyzed to a considerable extent by operator techniques without reference to special representations of the operators. The whole treatment of quantum-mechanical problems in terms of density matrices can thereby reflect the features of physical phenomena more directly and in closer correspondence to macroscopic methods than is otherwise possible.

This review proposes to familiarize the reader with the methods and the potentialities of quantum-mechanical treatments in terms of density matrices, utilizing a variety of examples. It starts by stating briefly the basic facts regarding the identification of states for atomic systems and regarding the properties of density matrices as developed by von Neumann (N27).* The treatment is then developed toward the systematic application of operator techniques. Most, if not all, of the factual material is well known, but the point of view departs in some respects from that of standard references (T38, N55).

The analysis of an experimental situation in Sec. 5 is designed to illustrate why partially polarized light should be described not as a mixture of different polarizations, but by parameters that define its observable properties without reference to fictitious models.

This review does not cover the field of quantum statistical mechanics, but it may clarify initial concepts and approaches to this field.

[1] The name "statistical matrix" is often used instead of "density matrix." The name density matrix itself relates to the correspondence between ρ and the distribution function $\rho(q_i,p_i)$ in the phase space of classical statistical mechanics. This correspondence has been developed by Wigner (W32), see Sec. 9.

* References are listed at the end of this article.

2. IDENTIFICATION OF STATES

(a) Pure States

Quantum mechanics deals usually with phenomena in which a maximum of information is available about the system under consideration. This maximal information is attained, for example, for a spinless particle whose state at a certain time is represented by a wave function $\psi(x,y,z)$, for a nonrelativistic spinning electron in a central field with a specified full set of quantum numbers (n,l,j,m), or for a light beam of intensity I, frequency ν, direction ω, and linear polarization A. States of maximal information are often called "pure states" or simply "states."

A pure state is characterized by the existence of an experiment that gives a result predictable with certainty when performed on a system in that state and in that state only. For example, linear polarization of a light beam in a given plane is characterized by 100% transmission of *each photon* through a suitably oriented Nicol prism; no other state of polarization is fully transmitted by the same prism. Filtration through a Nicol prism defines a state of polarization completely because beams thus filtered behave identically with respect to any other polarization analyzer.

An experiment that yields a unique predetermined result for a system in a given pure state can be designed to act as a filter which leaves the system undisturbed, like a Nicol prism traversed by light of the pertinent linear polarization. The experiment may then be repeated again and again on the same system, at least in principle, always with certainty as to its outcome. Pure states can, in fact, be "prepared" by subjecting systems to a filter-type experiment. The example of light polarization often proves particularly helpful to understand the relationships between quantum mechanical states and the experiments that characterize them; the relationships may be less obvious in other examples but are essentially the same.

An experiment that characterizes uniquely a pure state, as indicated above, and thus provides a maximum of information about it is called a "complete" experiment. The term "complete" may have a relative meaning, with reference to only a part of the variables of a system. For example, filtration through a Nicol prism is a complete experiment with regard to polarization, but has little relevance to the photon energy.

A pure state can then be identified by specifying the complete experiment that characterizes it uniquely. Mathematically one can construct a variety of Hermitian operators which have the given pure state as an eigenstate. An Hermitian operator which transforms the states of a physical system represents an observable.[2]

[2] This statement is believed valid for most applications in quantum mechanics but is subject to restrictions discussed in (WWW52).

Given such an operator, it proves possible to design, at least in principle, an experiment that constitutes a measurement of the corresponding observable. For example the state (n,l,j,m) of an electron bound in a central field is an eigenstate of the energy operator H corresponding to the eigenvalue E_{nlj} and also an eigenstate of the angular momentum operator J_z corresponding to the eigenvalue $m\hbar$. Experiments that measure energies and angular momenta are familiar; experiments to measure less familiar observables can nevertheless be designed.

When it is not convenient to identify a pure state by specifying the relevant complete experiment or its corresponding operator, the state may be identified as a linear superposition of eigenstates of any suitable complete set of operators. For example the wave function $\psi(x,y,z)$ of a particle describes its state as a superposition of all simultaneous eigenstates of the three position operators x, y and z, the amplitude of the eigenstate (x,y,z) being $\psi(x,y,z)$. The representation of a pure state, either as an eigenstate of a particular operator or as a superposition of eigenstates of another arbitrary operator, is usually called a "state vector" ψ. (It is also often called a wave function even when ψ is not expressed in terms of space coordinates.)

(b) General

Quantum-mechanical systems also occur for which no complete experiment gives a unique result predictable with certainty. For example, no polarization analyzer admits or rejects with certainty photons of partially polarized light. We say, loosely, that the information on such a system is less than a maximum, with reference to the lack of a complete experiment with a uniquely predetermined outcome. The state of the system is nevertheless fully identified by any data adequate to predict the (statistical) results of all conceivable observations on the system. Whether or not the predicted dispersion of these statistical results attains its theoretical minimum is irrelevant to the concept of state. Indeed "state" means whatever information is required about a specific system, in addition to physical laws, in order to predict its behavior in future experiments.

States that are not "pure" have been called "mixed" states because they can be described by the incoherent superposition of pure states. Incoherent superposition means, by definition, that to calculate the probability of finding a certain experimental result with a system in the mixed state one must first calculate the probability for each of the pure states and then take an average, attributing to each of the pure states an assigned "weight."

The concept of a nonpure state as a mixture of pure states originates from the early investigations of the connections between quantum mechanics and statistical mechanics. Classical statistical mechanics deals, ideally, with statistical mixtures (*ensembles*) of classic mechanical states, each of which is characterized by sharply defined values of all variables. The pure states of quantum mechanics were then visualized as analogues of the classic mechanical states and quantum-statistical mechanics was regarded as dealing with mixture of pure states.

However, the description of a nonpure state as the incoherent superposition of pure states is not unique. There is in general no reason, for example, why unpolarized light should be described as a mixture of two linear polarizations rather than of two circular ones.

We shall regard the state of a system, whether pure or not, as defined by its previous history, i.e., by the method of its preparation. Information on the preparation may be replaced by adequate experimentation on an ensemble of identically prepared systems. Fragmentary information, inadequate to make statistical predictions about future experiments, is often complemented by plausible assumptions. Examples given in Sec. 4 show how the information is represented by a density matrix. It will also be shown that lack of information about a variable can often be expressed as a statement that certain operators have expectation value zero. For example if the spin orientation of a particle is wholly unknown, the expectation value of each component of its angular momentum vanishes.

The point of view taken here differs to some extent from that which is often followed in standard treatments of statistical quantum mechanics, but the difference may be regarded as a matter of preference rather than of substance. To illustrate this difference, it may be recalled how Tolman (T38) analyzes the fluctuations in the experimental results obtained from a system in a given state. He considers first a subset of fluctuations that should be expected if the system were in any one of a number of pure states. Then he combines these subsets into a broader set according to the probability distribution that the system be in any of these pure states. In this paper we consider only the broader set of all fluctuations among the experimental results obtained with an ensemble of systems prepared according to identical specifications. We do not analyze these fluctuations into subsets when this analysis is not unique and does not correspond to observable characteristics of the situation. We regard as somewhat incidental whether a change in the preparation of the systems would reduce the fluctuations to a lower level or whether they already attain the minimum level set by quantum-mechanical laws. That is, we deal with a *single statistical ensemble* of quantum mechanical systems *prepared by identical procedures*, not with a statistical ensemble of quantum mechanical ensembles.[3] The terms "mean" or "average"

[3] The point of view of Elsasser (E37) is in some respects close to that adopted in this paper, but differs rather substantially in other respects, because it rests on the notion that "there exists no general *a priori* principle of how to collect samples which can be represented by the same statistical matrix" and that accordingly "the problem of determining the statistical matrix is essentially an indefinite one." Here we take a more positive attitude toward the determination of the statistical matrix.

will be always understood, in this review, to relate to the statistical expectation regarding such an ensemble.

3. DENSITY MATRIX

(a) Formulas for Mean Values

In the ordinary formalism of quantum mechanics, a pure state is identified by the coefficients c_n of the expansion of its state vector ψ into eigenvectors u_n of some complete set of operators,

$$\psi=\sum_n c_n u_n. \tag{3.1}$$

For a system in this state, an operator Q represented by a matrix $Q_{n'n}$ has the mean value

$$\langle Q\rangle=\sum_{n'n} Q_{n'n}c_{n'}^{*}c_n. \tag{3.2}$$

When a nonpure state is represented by the incoherent superposition of a number of pure states $\psi^{(i)}$ with statistical weights $p^{(i)}$, to each pure state corresponds a mean value $\langle Q\rangle_i$, and the mean value of Q for the incoherent superposition is given by the grand average

$$\langle Q\rangle=\sum_i p^{(i)}\langle Q\rangle_i=\sum_{nn'} Q_{n'n}\sum_i p^{(i)}c^{(i)}{}_{n'}{}^{*}c^{(i)}{}_n. \tag{3.3}$$

One defines then the density matrix as

$$\rho_{nn'}=\sum_i p^{(i)}c^{(i)}{}_{n'}{}^{*}c^{(i)}{}_n \tag{3.4}$$

so that (3.3) becomes

$$\langle Q\rangle=\sum_{n'n} Q_{n'n}\rho_{nn'}=\sum_{n'}(Q\rho)_{n'n'}=\mathrm{Tr}(Q\rho), \tag{3.5}$$

where TrA indicates the trace (i.e., the sum of the diagonal elements) of a matrix A.

(b) Definition

We regard the density matrix as defined by Eq. (3.5) rather than by (3.4). It represents a minimum set of input data which serves to calculate the mean value of any operator Q for a system prepared according to given specifications.[4] The information from which this set of data is derived is equivalent to a knowledge of the mean values of as many independent operators $Q^{(r)}$ as there are independent parameters in the matrix $\rho_{nn'}$. In fact the initial information on the state of the system is often conveniently expressed as a set of $\langle Q^{(r)}\rangle$ from which the $\rho_{nn'}$ are determined by solving a system of Eq. (3.5), one for each $\langle Q^{(r)}\rangle$. Equation (3.5) which gives $\langle Q\rangle$ for a generic Q in terms of the matrix ρ may be looked upon as a device to calculate a generic $\langle Q\rangle$ from advance knowledge of a special set of $\langle Q^{(r)}\rangle$'s.

(c) Limitations

Limitations on the matrix elements $\rho_{nn'}$ include the following:

(1) The condition that $\langle Q\rangle$ is real for every Hermitian operator Q, requires ρ to be *Hermitian* too,

$$\rho_{n'n}=\rho_{nn'}^{*}. \tag{3.6}$$

(2) The condition that the unit operator I has the mean value 1, requires

$$\mathrm{Tr}(I\rho)=\mathrm{Tr}(\rho)=\sum_n \rho_{nn}=1. \tag{3.7}$$

(3) The condition that every operator with non-negative eigenvalues (for example the operator $\delta_{kn}\delta_{kn'}$ with eigenvalues 1 and 0) has a non-negative mean value, requires ρ to be *positive definite*. That is, every diagonal element of ρ in any matrix representation must be non-negative,

$$\rho_{kk}\geqslant 0, \tag{3.8}$$

The mean value of the operator $\delta_{kn}\delta_{kn'}$, namely ρ_{kk}, represents the probability of finding the system by a suitable experiment in the pure state u_k. Accordingly, $\mathrm{Tr}\rho=\sum_k \rho_{kk}$ represents the total probability of finding the system in any one of a complete set of orthogonal states and the condition (3.7) constitutes an obvious requirement.

(4) The Hermitian matrix ρ may be reduced to the diagonal form

$$\rho_j\delta_{jj'}=\sum_{nn'} T_{jn}\rho_{nn'}T^{-1}{}_{n'j'} \tag{3.9}$$

by a unitary transformation T. The conditions (3.7) and (3.8) require that

$$\sum_j \rho_j^2\leq(\sum_j \rho_j)^2=[\mathrm{Tr}(\rho)]^2=1. \tag{3.10}$$

The $\sum_j \rho_j^2$ is the $\mathrm{Tr}(\rho^2)$, and therefore we have in general

$$\mathrm{Tr}(\rho^2)=\sum_{nn'}|\rho_{nn'}|^2\leq 1, \tag{3.11}$$

which limits the value of every single element of the density matrix.

(d) The Variations in Time

The variations in time of the density matrix are governed by the Schroedinger equation. Any unitary transformation S that changes the vector ψ representing a pure state into $S\psi$, also changes the matrix ρ representing a generic state into $S\rho S^{-1}$. The ordinary Schroedinger equation $\partial\psi/\partial t=-i\hbar^{-1}H\psi$ has, as a formal solution, the time variable transformation $\psi(t)=\exp(-i\hbar^{-1}Ht)\psi(0)$. The corresponding formulas for the density matrix are

$$\partial\rho/\partial t=-i\hbar^{-1}[H\rho-\rho H], \tag{3.12}$$

$$\rho(t)=\exp(-i\hbar^{-1}Ht)\rho(0)\exp(i\hbar^{-1}Ht). \tag{3.13}$$

The mean values of observables vary in time according to

$$\begin{aligned}\langle Q\rangle_t&=\mathrm{Tr}[Q\rho(t)]=\mathrm{Tr}[Q\exp(-i\hbar^{-1}Ht)\rho(0)\exp(i\hbar^{-1}Ht)]\\&=\mathrm{Tr}[\exp(i\hbar^{-1}Ht)Q\exp(-i\hbar^{-1}Ht)\rho(0)]\\&=\mathrm{Tr}[Q(t)\rho(0)].\end{aligned} \tag{3.14}$$

[4] Any prediction about the behavior of a system can be expressed as the mean value of a suitable operator Q. For example, the probability of a certain event is the mean value of an operator whose eigenvalues are 1 when the event occurs and 0 when it does not.

The transition from the Schroedinger representation ($Q=\text{const}$, $\rho=\rho(t)$) to the Heisenberg representation ($Q=Q(t)$, $\rho=\text{const}$) appears here as a trivial consequence of the identity $\text{Tr}(AB)=\text{Tr}(BA)$.

In the scheme where the Hamiltonian H is diagonal, with eigenstates v_m, (3.13) takes the form

$$\rho_{mm'}(t)=\rho_{mm'}(0)\exp[i\hbar^{-1}(E_{m'}-E_m)t]. \quad (3.15)$$

Notice that only energy differences and the corresponding frequencies $\hbar^{-1}(E_{m'}-E_m)$ appear here, to the exclusion of absolute energies which are indeed not observable.

(e) The Diagonal Representation

The diagonal representation (3.9) of the density matrix is equivalent to

$$\rho_{nn'}=\sum_j \rho_j T_{jn'}{}^* T_{jn}. \quad (3.16)$$

Comparison of (3.16) with (3.4) shows that a generic state can be represented by the incoherent superposition of orthogonal pure states ψ_j with statistical weights ρ_j, the state vectors ψ_j being identified by their components which are the elements of the transformation matrix T_{jn}. This representation is *not unique* in general. That is, an infinity of different representations (3.4) can be constructed for a given density matrix unless: (a) one stipulates that the states $\psi^{(i)}$ represented by the $c_n{}^{(i)}$'s are mutually orthogonal, like the ψ_j's, and (b) no two nonzero eigenvalues ρ_j are equal. Examples of the relevance of these conditions will be given in the following sections.

(f) A Pure State

A pure state is represented by a density matrix with one eigenvalue equal to 1 and all others equal to 0. (The eigenvalues of the matrix are invariant under unitary transformations). Equation (3.16) reduces then to $\rho_{nn'}=T_{1n'}{}^*T_{1n}$. A density matrix $\rho_{nn'}$ can be factorized in this manner when, and only when, all minors of its determinant vanish.

(g) The Number of Independent Parameters

The number of independent parameters that identify a density matrix depends on its number N of rows and columns. This is the number of orthogonal pure states over which the $\sum_n$ extends in (3.1) and in the subsequent formulas. This number may be infinite but is often finite when one considers only a particular property of a system, such as the spin orientation of a nucleus.

A matrix with N rows and columns has N^2 elements. The condition (3.6), that ρ be Hermitian, restricts the number of independent real parameters in the N^2 complex elements to N^2, and (3.7) restricts it further to N^2-1. These parameters are further limited in their range of variation by inequalities like (3.8) and (3.11), but are nevertheless independent. Therefore *it takes*, in general, *N^2-1 separate measurements*, or equivalent data, *to identify the state* of a system which possesses N independent pure states. For example, a nucleus of Na^{23} with spin $\frac{3}{2}$ has 4 independent pure states of spin orientation and the description of a specific state of orientation requires in general 15 independent data.[5]

Note, for comparison, that a pure state is identified by the N complex coefficients c_n of (3.1). The number of meaningful independent real parameters in the c_n's is reduced from $2N$ to $2N-1$ by the normalization condition $\sum_n c_n{}^2=1$, and further to $2N-2$ because the phase of the state vector ψ is arbitrary and physically meaningless.

(h) Interacting Systems

The joint state of two interacting systems a and b is represented by a density matrix $\rho^{(ab)}$ of which each row or column is labelled by two indices (m,n) corresponding to eigenstates u_m and v_n of the separate systems.[6] An operator $Q^{(a)}$ of the system a may be treated as an operator of the whole system when multiplied by the unit operator of b, $1^{(b)}=\delta_{n'n}$. The mean value of $Q^{(a)}$ is then

$$\begin{aligned}\langle Q^{(a)}\rangle &= \text{Tr}(Q^{(a)}1^{(b)}\rho^{(ab)})\\ &= \sum_{mm'} Q_{m'm}{}^{(a)}[\sum_{nn'}\delta_{n'n}\rho_{mn,m'n'}{}^{(ab)}]\\ &= \text{Tr}_a(Q^{(a)}\rho^{(a)}), \quad (3.17)\end{aligned}$$

where

$$\rho_{mm'}{}^{(a)}=\sum_n \rho_{mn,m'n}{}^{(ab)}=[\text{Tr}_b(\rho^{(ab)})]_{mm'} \quad (3.18)$$

represents the information on a alone contained in $\rho^{(ab)}$.

The states of the two systems are *uncorrelated* when

$$\langle Q^{(a)}Q^{(b)}\rangle=\langle Q^{(a)}\rangle\langle Q^{(b)}\rangle \quad (3.19)$$

for all pairs $Q^{(a)}Q^{(b)}$ which requires the joint density matrix to be a product,

$$\rho_{mn,m'n'}{}^{(ab)}=\rho_{mm'}{}^{(a)}\rho_{nn'}{}^{(b)}. \quad (3.20)$$

4. EXAMPLES

(a) Orientation of Spin-1/2 Particles

This spin orientation is represented by a density matrix with two rows and columns, corresponding to

[5] The clause "in general" refers to the possibility that a few initial measurements turn out to be compatible only with certain special values of the remaining parameters. In this event further measurements of these parameters would be unnecessary. For example, if a first experiment on an assembly of Na^{23} nuclei utilizes a Stern-Gerlach inhomogeneous magnet and if every nucleus is found to follow the path corresponding to the magnetic quantum number $m=\frac{1}{2}$, the state is already fully identified.

[6] The two systems may consist of two portions of matter, of two different groups of particles (like nuclei and electrons of the same portion of matter) or of different characteristics (like spin and orbital motion) of the same particle. One can regard a system as an aggregate of two separate systems whenever there exists a complete set of operators O, pertaining to the whole system, that is the sum of two sets of operators $O^{(a)}$ and $O^{(b)}$, such that $O^{(b)}$ does not operate on the eigenstates u_m of $O^{(a)}$, and $O^{(a)}$ does not operate on the v_n's of $O^{(b)}$.

two pure states of opposite spin orientation (e.g., up and down). In a generic state the degree and direction of spin orientation are indicated by the magnitude and direction of the vector

$$\mathbf{P}=\langle\boldsymbol{\sigma}\rangle=\mathrm{Tr}(\rho\boldsymbol{\sigma}) \tag{4.1}$$

whose components are the mean values of the operators represented by the three Pauli matrices σ_x, σ_y, and σ_z. A pure state with definite spin orientation has $P=1$, a state of random orientation has $P=0$. For particles with magnetic moment μ in a magnetic field $\mathbf{H}$, under conditions of paramagnetic polarization, we have $\mathbf{P}=\mu\mathbf{H}/\frac{1}{2}kT$.

Knowledge of $\mathbf{P}$ is sufficient to identify the density matrix, considering: (1) that any 2×2 Hermitian matrix can be represented as a linear combination of σ_x, σ_y, σ_z and of the unit matrix $\mathcal{I}$, (2) that the Pauli matrices have the properties $\mathrm{Tr}\sigma_i=0$, $\mathrm{Tr}(\sigma_i\sigma_k)=2\delta_{ik}$, (3) that $\mathrm{Tr}\rho=1$ and $\mathrm{Tr}\mathcal{I}=2$ require the coefficient of $\mathcal{I}$ to be $\frac{1}{2}$. We have, then, (TG49)

$$\rho=\tfrac{1}{2}(\mathcal{I}+P_x\sigma_x+P_y\sigma_y+P_z\sigma_z)=\tfrac{1}{2}(\mathcal{I}+\mathbf{P}\cdot\boldsymbol{\sigma})$$
$$=\tfrac{1}{2}\begin{vmatrix} 1+P_z & P_x-iP_y \\ P_x+iP_y & 1-P_z \end{vmatrix}. \tag{4.2}$$

The representation (4.2) leads to a simple treatment of the Larmor precession of spin orientation in a magnetic field. The magnetic moment $\boldsymbol{\mu}$ is represented quantum-mechanically by the operator $\gamma(\frac{1}{2})\hbar\boldsymbol{\sigma}$, where γ is the gyromagnetic ratio, so that the Hamiltonian in a magnetic field $\mathbf{H}$ is $-\gamma(\frac{1}{2})\hbar\boldsymbol{\sigma}\cdot\mathbf{H}$. The Schroedinger equation (3.12) is then

$$\partial\rho/\partial t=\tfrac{1}{2}\partial\mathbf{P}/\partial t\cdot\boldsymbol{\sigma}=i\tfrac{1}{4}\gamma[\mathbf{H}\cdot\boldsymbol{\sigma}\;\mathbf{P}\cdot\boldsymbol{\sigma}-\mathbf{P}\cdot\boldsymbol{\sigma}\;\mathbf{H}\cdot\boldsymbol{\sigma}]$$
$$=-\tfrac{1}{2}\gamma\mathbf{H}\times\mathbf{P}\cdot\boldsymbol{\sigma}, \tag{4.3}$$

and thus reduces to the classical equation

$$\partial\mathbf{P}/\partial t=-\gamma\mathbf{H}\times\mathbf{P}. \tag{4.4}$$

The extension of this equation to spin values $j>\frac{1}{2}$ is indicated below, the extension to include the paramagnetic relaxation is given in Sec. 11g.

The response of a particle detector, whose efficiency depends on spin orientation, and which therefore serves as a polarization analyzer, may be represented by an operator in a form analogous to (4.2). Maximum and minimum efficiency, ϵ_M and ϵ_m, correspond necessarily, for spin $\frac{1}{2}$, to particles with opposite spin orientations indicated by unit vectors $\mathbf{Q}$ and $-\mathbf{Q}$. The detector is then represented by an operator D with the spin orientation eigenstates $\mathbf{Q}$ and $-\mathbf{Q}$ and the eigenvalues ϵ_M and ϵ_m. When the z axis is parallel to $\mathbf{Q}$, the matrix of this operator is

$$D=\begin{vmatrix} \epsilon_M & 0 \\ 0 & \epsilon_m \end{vmatrix}, \tag{4.5}$$

and for a generic orientation of axes it is

$$D=\tfrac{1}{2}[(\epsilon_M+\epsilon_m)\mathcal{I}+(\epsilon_M-\epsilon_m)\mathbf{Q}\cdot\boldsymbol{\sigma}]. \tag{4.6}$$

The probability of response of this detector to particles with the density matrix (4.2) is the mean value

$$\langle D\rangle=\mathrm{Tr}(\rho D)=\tfrac{1}{2}[(\epsilon_M+\epsilon_m)+(\epsilon_M-\epsilon_m)\mathbf{P}\cdot\mathbf{Q}]. \tag{4.7}$$

(b) Orientation of Spin-j Particles

The density matrix of spin orientation has $2j+1$ rows and columns. For $j>\frac{1}{2}$ the representation (4.2) can be generalized to include, besides the unit matrix and terms depending on dipole polarization, additional terms which depend on multipole polarization (see end of Sec. 6). This expansion of ρ constitutes a "reduction" in the sense of tensor algebra (FR57, Sec. 19) and the coefficients of the operators may be called the "reduced elements" of ρ.[7] The application of the expansion to collision and disintegration processes will be outlined in Sec. 11 a.

The definition (4.1) of the spin orientation vector is replaced for a generic j by

$$\mathbf{P}=\langle\mathbf{J}\rangle/j\hbar, \tag{4.8}$$

where $\mathbf{J}$ indicates the angular momentum operator. With this definition, the changes of spin orientation induced by a magnetic field $\mathbf{H}$ are described by (4.4) for all values of j.[8,9] The multipole parameters of spin orientation remain constant, under the influence of $\mathbf{H}$, provided they are expressed in a coordinate system which precesses with $\mathbf{P}$.

(c) Polarization of Electromagnetic Radiation

Light polarization is represented by a density matrix with two rows and columns, corresponding to two opposite polarizations, e.g. to linear polarizations indicated by orthogonal unit vectors $\mathbf{A}_1$ and $\mathbf{A}_2$. Because of mathematical analogy to the density matrix for the orientation of spin-$\frac{1}{2}$ particles, the density matrix of light polarization can be represented in the form (4.2), but the indices x, y, and z no longer correspond to directions of the physical space. They relate to a mathematical representation of polarizations in a 3-dimen-

---

[7] These coefficients have also been called "statistical tensors" (F51a) and "state multipoles" (F53).

[8] This result is derived from the Schroedinger Eq. (3.12), taking into account that: (1) the magnetic moment $\boldsymbol{\mu}$ is represented by the operator $\gamma\mathbf{J}$, (2) therefore, the Hamiltonian is $-\gamma\mathbf{J}\cdot\mathbf{H}$, (3) the definition (4.8) requires the dipole term in the expansion of ρ to have the form $3\mathbf{P}\cdot\mathbf{J}[j(j+1)(2j+1)\hbar]^{-1}$, (4) the commutator $[\mathbf{J}\cdot\mathbf{H},\mathbf{P}\cdot\mathbf{J}]$ equals $i\hbar\mathbf{H}\times\mathbf{P}\cdot\mathbf{J}$.

[9] The history of (4.4) has some interest. The treatment of spin precession by ordinary quantum-mechanical methods is very complicated (BR45). Bloch pointed out, in his paper on nuclear induction (B46), that (4.4) can be established without solving the Schroedinger equation, because the quantum-mechanical mean value of any quantity follows in its time dependence exactly the classical equations of motion. Later Wangsness and Bloch (WB53) derived (4.4) from the Schroedinger equation by a density matrix procedure equivalent to that indicated here. The link between the classical and the Schroedinger equation emerges clearly when ρ is expressed in terms of mean values, as in (4.2), so that the Schroedinger equation for ρ reduces to an equation of motion for mean values. A general formulation of this reduction will be given in Sec. 6.

sional space, called "Poincaré representation," in which the z axis corresponds to linear polarization along $\mathbf{A}_1$, negative z to polarization along $\mathbf{A}_2$, positive x to linear polarization at 45° between $\mathbf{A}_1$ and $\mathbf{A}_2$ and positive y to circular polarization rotating from $\mathbf{A}_1$ toward $\mathbf{A}_2$. A unitary transformation of the base polarizations $\mathbf{A}_1$ and $\mathbf{A}_2$ is represented by a rotation of Cartesian axes in the representative space.

This representation of the density matrix has proved convenient (F49, F54a), in the same manner as the corresponding representation for spinning particles. The form (4.2) of the density matrix is normalized to provide the input for the calculation of elementary processes involving one photon at a time. For macroscopic processes, one multiplies (4.2) simply by the intensity I of a radiation beam. The quantities I, IP_x, IP_y, IP_z constitute a set of Stokes parameters which characterizes simultaneously the intensity and polarization state of the beam (S52, F49, FMcD51).

(d) Joint Polarization of Photon Pairs

In the positron annihilation by two-photon process, which was mentioned in Sec. 1, the emitted γ rays display a polarization only when the two photons of a pair are analyzed simultaneously but not when the photons are observed singly. The experimental results can be described in terms of the mean value $\langle D^{(A)}D^{(B)}\rangle$, where the operators D represent polarization analyzers serving to detect two photons A and B in coincidence. The matrix of $D^{(A)}$ or $D^{(B)}$ is given by (4.6), where $\mathbf{Q}$ indicates now a vector of the Poincaré representation of polarizations. The usual statement that "the photons of a pair have opposite polarizations" means actually that, for perfect analyzers with $\epsilon_m=0$, one would find $\langle D^{(A)}D^{(B)}\rangle=(1/4)\epsilon_M{}^{(A)}\epsilon_M{}^{(B)}(1-\mathbf{Q}^{(A)}\cdot\mathbf{Q}^{(B)})$, which vanishes for $\mathbf{Q}^{(A)}=\mathbf{Q}^{(B)}$ and equals $(\frac{1}{2})\epsilon_M{}^{(A)}\epsilon_M{}^{(B)}$ for opposite analyzer settings $\mathbf{Q}^{(A)}=-\mathbf{Q}^{(B)}$, whether $\mathbf{Q}^{(A)}$ corresponds to linear, circular, or elliptical polarization. In a realistic application with imperfect analyzers one finds

$$\langle D^{(A)}D^{(B)}\rangle=\tfrac{1}{4}[(\epsilon_M+\epsilon_m)^{(A)}(\epsilon_M+\epsilon_m)^{(B)} -(\epsilon_M-\epsilon_m)^{(A)}(\epsilon_M-\epsilon_m)^{(B)}\mathbf{Q}^{(A)}\cdot\mathbf{Q}^{(B)}]. \quad (4.9)$$

The joint state of polarization of the photon pairs is represented by a density matrix $\rho^{(AB)}$ with 4 rows and columns, which can always be expressed as a linear combination of the 16 matrices $\mathit{1}^{(A)}\mathit{1}^{(B)}$, $\mathit{1}^{(A)}\sigma_x{}^{(B)}$, $\cdots\sigma_z{}^{(A)}\sigma_z{}^{(B)}$. The requirement that $\mathrm{Tr}(\rho^{(AB)}D^{(A)}D^{(B)})$ be given by (4.9) for all choices of ϵ and $\mathbf{Q}$ implies that

$$\rho^{(AB)}=\tfrac{c}{4}(\mathit{1}^{(A)}\mathit{1}^{(B)}-\boldsymbol{\sigma}^{(A)}\cdot\boldsymbol{\sigma}^{(B)}). \quad (4.10)$$

This matrix represents the situation in very condensed form, without reference to any specific type of polarization. The negative sign corresponds to the statement that the "photons have *opposite* polarization." The density matrix for a single photon is, according to (3.18), $\rho^{(A)}=\mathrm{Tr}_B\rho^{(AB)}=(\frac{1}{2})\mathit{1}^{(A)}$ and indicates natural polarization as expected.

The joint density matrix (4.10) can be derived, of course, by a theoretical calculation of the annihilation process.

(e) Steady States (Constant in Time)

When the state of a system does not vary in the course of time, the density matrix is constant. According to (3.15) we have then, in the scheme of energy eigenstates,

$$\rho_{mm'}=0 \quad \text{for} \quad E_m\neq E_{m'}. \quad (4.11)$$

Any operator Q of the system may be split into two parts, $Q=Q^{(c)}+Q^{(v)}$, one constant and one variable, by defining

$$Q_{mm'}{}^{(c)}=Q_{mm'}\delta_{E_mE_{m'}}, \quad Q_{mm'}{}^{(v)}=Q_{mm'}(1-\delta_{E_mE_{m'}}). \quad (4.12)$$

A steady state may then be characterized by the property

$$\langle Q^{(v)}\rangle=0 \quad \text{for all } Q. \quad (4.13)$$

In the Heisenberg representation formalism, $Q^{(v)}$ may be defined by $Q^{(v)}(t)=Q(t)-Q(0)$.

(f) States with Space Invariances

Space invariance may be treated like time invariance. For example if the state of a system is invariant under translations along the x axis, its density matrix has a property analogous to (4.11) in the scheme of momentum eigenstates, namely

$$\rho_{p_xp_{x'}}=0 \quad \text{for} \quad p_x\neq p_{x'}. \quad (4.14)$$

States invariant under space rotation, i.e., states of random orientation and spherical symmetry, are very important. Here the relevant scheme is that of eigenstates of the squared angular momentum $\mathbf{J}^2$ and of one of its components, e.g., J_z. Spherical symmetry is represented by

$$\rho_{\alpha jm,\alpha' j'm'}=0 \quad \text{for} \quad j\neq j' \quad \text{or} \quad m\neq m', \quad (4.15)$$

where j and m relate, as usual, to the eigenvalues of $\mathbf{J}^2$ and J_z; α is an additional quantum number relating to the eigenvalues of any additional operators that commute with $\mathbf{J}^2$, J_z and among themselves.

In the more restricted case of axial symmetry only, we have

$$\rho_{\alpha jm,\alpha' j'm'}=0 \quad \text{for} \quad m\neq m'. \quad (4.16)$$

A nucleus that had initially random spin orientation and has absorbed, or interacted with, a beam of unpolarized radiation has axial symmetry about the direction of the beam.[10] Any operator Q can be split, as in (4.12), into a

[10] This property is fundamental for the theory of angular correlations. It was first stated tentatively, in a different form, by Falkoff and Uhlenbeck (FU50) and then proved by various authors (L50, S50, L51, TG51), through arguments equivalent to the elementary considerations which lead to (4.16).

part $Q^{(s)}$ having a symmetry property of interest, e.g. invariance under space rotation, and a residual part $Q^{(\text{non-}s)}$ which is not invariant. The mean value $\langle Q^{(\text{non-}s)}\rangle$ vanishes for symmetric states.

(g) States of Thermal Equilibrium

Statistical mechanics shows that the state of a system at a temperature T is represented by the incoherent superposition of eigenstates of energy E_m with weights proportional to the Boltzmann distribution factor $\exp(-E_m/kT)$. In order that the sum of the weights for all eigenstates equal 1, the weight of each state must equal the factor $\exp(-E_m/kT)$ divided by the "sum of states" $Z(T)=\sum_m \exp(-E_m/kT)$. That is, the density matrix is diagonal in the scheme of energy eigenstates and is given by

$$\rho_{mm'}=\frac{e^{-E_m/kT}}{Z(T)}\delta_{mm'}, \tag{4.17}$$

in this scheme and by

$$\rho=\frac{e^{-H/kT}}{Z(T)}=\frac{e^{-H/kT}}{\mathrm{Tr}(e^{-H/kT})} \tag{4.18}$$

in operator notation.

Because ρ is a function of the Hamiltonian operator only, and this function can be expanded in series, the mean value of any operator Q can be expressed in terms of the traces $\mathrm{Tr}(QH^n)$. In particular we have

$$\langle Q^{(i)}\rangle=0 \tag{4.19}$$

for the whole class of operators $Q^{(i)}$ such that $\mathrm{Tr}(Q^{(i)}H^n)=0$ for all n, i.e. such that $\mathrm{Tr}[Q^{(i)}f(H)]=0$ for all functions of H. The derivation of (4.18) can be inverted to show that, whenever (4.19) holds, ρ is a function of H only.

The functional dependence of ρ on H can be identified by the mean values $\langle H\rangle$, $\langle H^2\rangle\cdots\langle H^n\rangle\cdots$, which are given by well known formulas of statistical mechanics in terms of the derivatives of $Z(T)$, e.g., $\langle H\rangle=-d\ln Z/d(1/kT)$, $\langle H^2\rangle=\langle H\rangle^2+[d/d(1/kT)]^2\ln Z$, etc.

The derivation of (4.18) through statistical mechanics is, then, equivalent to showing that, for a system in contact with a thermal reservoir, the mean values of all operators $Q^{(i)}$ tend to zero in the course of time and the $\langle H^n\rangle$ approach the values predicted by thermodynamics. These results should follow from the Wangsness-Bloch equation discussed in Sec. 11g.

5. ANALYSIS OF LIGHT EMITTED BY ATOMS

As a further illustration we discuss here the polarization of light emitted by atoms under controlled conditions in an ideal experimental arrangement (F49) (Fig. 1). The emission takes place in a chamber C on which an optical detector and polarization analyzer are collimated from a variable direction. The atoms enter

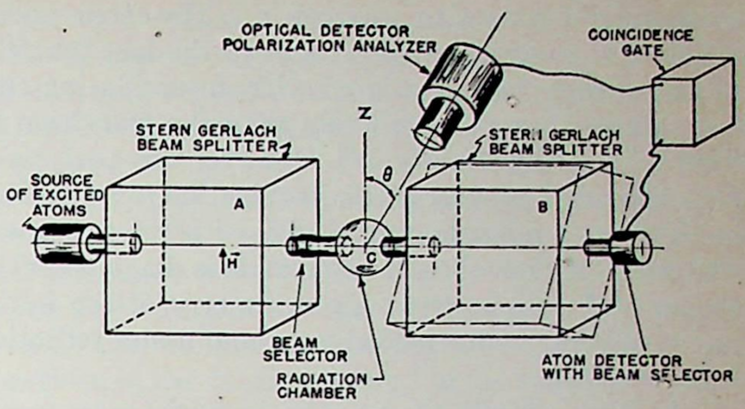


FIG. 1. Block diagram of an ideal experiment.

the chamber as a molecular beam, preanalyzed by a Stern-Gerlach device A to select atoms with a specified magnetic quantum number m, referred to an axis z parallel to the magnetic field in A. To be specific, we consider an optical transition from a p state ($l=1$) to a d state ($l'=2$) and assume that the Stern-Gerlach analyzer A has removed all atoms with $m\neq 0$. In the ordinary language of quantum mechanics we say, then, that the transition will lead from ($l=1$, $m=0$) to ($l'=2$, $m'=0,\pm 1$).

The angular distribution of intensity and polarization of the light emitted by these atoms can be calculated, for example, by considering the light as the incoherent superposition of three components emitted respectively in transitions to $m'=1$, 0 and -1. Alternately, one can derive by a density matrix treatment a general formula (FR57, 19.6) for the intensity and polarization distribution of radiation emitted in transitions to a state whose orientation is not observed.

The result can be expressed as a density matrix for the polarization, in the form (4.2), multiplied by an intensity distribution. Both the density matrix and the intensity distribution are functions only of the angle θ between the direction of emission and the z axis, since the state of atoms filtered by A has cylindrical symmetry about this axis if the atom velocity is disregarded. The result is

$$I(\theta)\rho\propto(1+\tfrac{1}{6}\sin^2\theta)\left[I+\frac{\sin^2\theta}{6+\sin^2\theta}\sigma_\zeta\right], \tag{5.1}$$

where the subscript ζ indicates the direction in the representative Poincaré space corresponding to linear polarization in a plane containing the z axis. In explicit matrix notation with the first row and column referring to this polarization and the second to linear polarization orthogonal to z, we have

$$I(\theta)\rho\propto\begin{vmatrix}1+(4/3)\sin^2\theta & 0\\ 0 & 1\end{vmatrix}. \tag{5.2}$$

Because the matrix is diagonal in this scheme, the light can be regarded as the incoherent superposition of two beams, one polarized linearly in a vertical plane, with intensity $\propto 1+(4/3)\sin^2\theta$ and one polarized in a

horizontal direction with intensity $\propto 1$. The excess polarization in the vertical plane relates to the fact that the transition to $m'=0$, associated with an oscillating current parallel to z, is more likely than the transition to either $m'=\pm 1$ by a factor 4/3. However this analysis of the polarization provides little justification for regarding the light as "a mixture of photons" with the two linear polarizations. That the density matrix is diagonal in the scheme (5.2) follows from the symmetry of the initial state $m=0$ under rotations about z and under reflection on the xy plane.

If the device A selects atoms in the initial state $m=1$, instead of $m=0$, the symmetry under reflection on the xy plane drops out and the density matrix of the polarization is more complicated. We have here

$$I(\theta)\rho \propto \left(1-\frac{1}{14}\sin^2\theta\right)\left[\mathbf{1}+\frac{10\cos\theta\,\sigma_\eta-\sin^2\theta\,\sigma_\zeta}{14-\sin^2\theta}\right]$$
$$=\left(1-\frac{1}{14}\sin^2\theta\right)\left[\mathbf{1}+\frac{(100\cos^2\theta+\sin^4\theta)^{\frac{1}{2}}}{14-\sin^2\theta}\times(\cos\alpha\,\sigma_\eta-\sin\alpha\,\sigma_\zeta)\right], \quad (5.3)$$

where the subscript η refers to circular polarization and the angle $\alpha=\mathrm{arctg}(\sin^2\theta/10\cos\theta)$ indicates the type of partial polarization, which varies from left circular to elliptical to linear and again to elliptical and right circular as θ varies from 0° to 180°.

Consider now another Stern-Gerlach device B added at the exit of the radiation chamber C, which can select atoms having a specified quantum number m' *after* the light emission process. These atoms are to be detected at the exit of B and the detector output can be used to "gate" the light analyzer device. The light will then appear to be completely polarized. If the magnetic fields in B and A are parallel and A selects $m=0$, atoms will pass through B in coincidence with radiation emission for $m'=1$, 0 or -1. For $m'=0$ the polarization is linear, in a vertical plane; for $m'=\pm 1$ it is circular, elliptic or linear depending on the direction of observation. The intensity and polarization distribution for each m' is given by

$$I(\theta)\rho \propto \tfrac{2}{3}\sin^2\theta(\mathbf{1}+\sigma_\zeta), \quad \text{for } m'=0$$
$$I(\theta)\rho \propto \tfrac{1}{4}(1+\cos^2\theta)\left(\mathbf{1}-\frac{\sin^2\theta\sigma_\zeta\mp 2\cos\theta\sigma_\eta}{1+\cos^2\theta}\right), \quad (5.4)$$
$$\text{for } m'=\pm 1.$$

The sum of the three intensity and polarization distributions yields (5.1).

The elementary treatment leads one to visualize the partially polarized light represented by (5.1) as a mixture of three kinds of photons polarized, respectively, according to (5.4), simply because it appears natural to classify the final state according to the same axis of quantization as the initial state. (Notice that we have here *three* incoherent components, whereas the diagonal matrix (5.2) suggests only two.) However, this representation does not rest on any intrinsic property of the emitted light. Indeed, *if the analyzer B is tilted* about the axis of the atomic beam, so that the final quantum number m' relates to an axis oblique to z, the beam of atoms after radiation will be resolved into *five* components, with $m'=0$, ± 1, ± 2. Each of these components is associated with a light emission fully polarized, so that the light is in fact resolved into *five* incoherent polarized components, when the analyzer is gated by the output of B. Each of the five components is represented by a matrix $I(\theta)\rho$ and the sum of these matrices equals (5.1).

In summary, light is completely polarized only when maximum information regarding the state of orientation of the atoms which have emitted it (or have interacted with it) is provided, and actually taken into account. The information is required regarding the orientation both before and after the process. Incomplete information leads to partial polarization. Different representations of partial polarization as the incoherent superposition of polarized components correspond to different conceivable methods of completing the information on the atoms. None of these alternate representations appears intrinsically relevant—apart from analytical convenience—unless the complete information on the atom is actually collected and correlated with the polarization analysis.

An experimental analysis, equivalent to that indicated here for light polarization, can be developed for the orientation of particle spins or for any other quantum-mechanical variable.

We have dealt here specifically with the information or lack of information on the orientation of the atoms *after* light emission, that is, with a follow-up analysis. One could as well have considered alternate types of pre-analysis. For example one could arrange the selective device A so that it first splits up the atom beam components with different m and then refocuses them together on the entrance of the radiation chamber. It is then possible to select one or the other m value at will by stopping the other beam components with a diaphragm, or else to allow two or more beam components to radiate together.

6. EXPANSION IN ORTHOGONAL OPERATORS

The density matrix of the orientation of spin-($\frac{1}{2}$) particles is decomposed by (4.2) into a sum of standard matrices $\mathbf{1}$ and $\boldsymbol{\sigma}$. This decomposition was found convenient. It replaces, in effect, the elements of the density matrix with the coefficients of the standard matrices, namely 1, P_x, P_y, and P_z in (4.2), which are directly observable parameters being the mean values of physical quantities. Indeed (4.4) constitutes an equation

of motion of the spin expressed entirely in terms of the observable vector parameter **P**.

This representation can be generalized. The operators $\mathcal{I}$, σ_x, σ_y, and σ_z are orthogonal, in the sense that the product of any two of these operators has trace zero. For a generic system one may consider a set of operators U_i which are Hermitian, except when otherwise noted, and obey the orthogonality condition

$$\mathrm{Tr}(U_i U_k) = C\delta_{ik}. \tag{6.1}$$

C represents a normalization factor which will be taken as 1, for convenience, unless otherwise noted. For systems whose complete operators have N orthogonal eigenstates, so that ρ has N rows and columns, N^2 linearly independent operators constitute a complete orthogonal set of base operators.

Each operator Q of the system can be expanded into a sum of U_i's,

$$Q = \sum_i \mathrm{Tr}(QU_i)U_i = \sum_i Q_i U_i. \tag{6.2}$$

In particular the density matrix is represented by

$$\rho = \sum_i \mathrm{Tr}(\rho U_i)U_i = \sum_i \langle U_i\rangle U_i = \sum_i \rho_i U_i. \tag{6.3}$$

The state is thereby identified by the parameters ρ_i which are the mean values $\langle U_i\rangle$ of the base operators. The mean value of any operator Q is a linear combination of the ρ_i's,

$$\langle Q\rangle = \mathrm{Tr}(Q\rho) = \sum_i Q_i \rho_i. \tag{6.4}$$

Suitable choice of the U_i's serves to eliminate irrelevant variables, much like the choice of coordinates does in classical physics. The initial state of a system is often identified by the fact that certain types of operators have mean value zero, as seen in the examples of Sec. 4. It is then convenient to include these operators among the U_i's. With regard to the final state of the system one is often interested only in predicting the results of certain experiments, i.e. the mean values of only a few operators. If these operators are included in the U_i's, then only a few of the ρ_i's have to be calculated. This approach is most effective, of course, when the equations of motion interlink only some subsets of the ρ_i's.

It is usually convenient that one of the base operators be the unit operator, to within a normalization constant. This operator will be indicated as

$$U_0 = [\mathrm{Tr}(\mathcal{I})]^{-\frac{1}{2}}\mathcal{I}, \tag{6.5}$$

where the normalization is such that $C=1$ in (6.1). The operator U_0 commutes with all other operators, and its mean value

$$\rho_0 = \langle U_0\rangle = [\mathrm{Tr}(\mathcal{I})]^{-\frac{1}{2}} \tag{6.6}$$

is the same for all states of the system, owing to the normalization condition $\mathrm{Tr}\{\mathcal{I}\rho\}=1$. A state is then identified by the N^2-1 parameters ρ_i with $i\neq 0$. The condition (6.1) requires that

$$\mathrm{Tr}\{U_0U_i\} = \mathrm{Tr}\{\mathcal{I}U_i\} = \mathrm{Tr}\{U_i\} = 0, \quad \text{for} \quad i\neq 0, \tag{6.7}$$

i.e., that the sum of the eigenvalues of U_i vanish for $i\neq 0$. States of random spin orientation or of zero polarization, i.e., states of minimum information, are represented by a density matrix proportional to $\mathcal{I}$ ("equiprobability") and, therefore, by zero values of ρ_i for $i\neq 0$. Departures of ρ_i from zero represent positive elements of information about the state.

When there are N rows and columns, $U_0 = N^{-\frac{1}{2}}\mathcal{I}$. When N is infinitely large, the situation is somewhat akin to that encountered in the normalization of wave functions in the continuum and in the application of the Dirac δ function.

The Schroedinger equation (3.12) takes the form of a system of linear equations among the mean values $\rho_i = \langle U_i\rangle$. Replacing ρ with $\sum_k \rho_k U_k$ on the right of (3.12), multiplying (3.12) by U_i and taking the trace one obtains

$$\begin{aligned}\partial\rho_i/\partial t &= -i\hbar^{-1}\sum_k \mathrm{Tr}\{U_i[H,U_k]\}\rho_k\\ &= i\hbar^{-1}\sum_k \mathrm{Tr}\{H[U_i,U_k]\}\rho_k,\end{aligned} \tag{6.8}$$

that is

$$\partial\rho_i/\partial t = \sum_k \Omega_{ik}\rho_k, \tag{6.9}$$

where Ω_{ik} is an antisymmetric matrix. The set of commutators $[U_i,U_k]$ can again be expanded into a sum of base operators

$$i[U_i,U_k] = \sum_p c_{ik}{}^p U_p, \quad c_{ik}{}^p = \mathrm{Tr}\{U_p i[U_i,U_k]\}, \tag{6.10}$$

where the coefficients $c_{ik}{}^p$, antisymmetric in (i,k) and real, are an important property of the set of base operators.

The matrix Ω_{ik} of the system of Eqs. (6.9) is therefore real,

$$\Omega_{ik} = \sum_p c_{ik}{}^p\, \mathrm{Tr}\{\hbar^{-1}HU_p\} = -\Omega_{ki}, \tag{6.11}$$

and represents an infinitesimal orthogonal transformation. Notice that $\Omega_{i0}=\Omega_{0k}=0$ since all U_i's commute with U_0. The eigenvalues of Ω are imaginary and equal to i times the beat frequencies $\hbar^{-1}(E_{m'}-E_m)$ among the eigenvalues E_m of H, as seen in (3.15).

The operators U_i need not be Hermitian. If they are not, one must consider also their adjoint, or Hermitian conjugate, operators $U_i^+ = \tilde{U}_i^*$ (where $\sim$ indicates transposition and * complex conjugation). Equations (6.1) to (6.4) are then replaced by

$$\mathrm{Tr}\{U_iU_k^+\} = \delta_{ik}, \tag{6.12}$$

$$Q = \sum_i \mathrm{Tr}\{QU_i^+\}U_i = \sum_i Q_i^+ U_i = \sum_i Q_i U_i^+, \tag{6.13}$$

$$\rho = \sum_i \mathrm{Tr}\{\rho U_i^+\}U_i = \sum_i \rho_i^+ U_i = \sum_i \rho_i U_i^+, \tag{6.14}$$

$$\langle Q\rangle = \sum_i Q_i^+\rho_i = \sum_i Q_i\rho_i^+, \tag{6.15}$$

and (6.10) by

$$\begin{aligned} i[U_i,U_k^+] &= \sum_p c_{ik}{}^p U_p,\\ c_{ik}{}^p &= \mathrm{Tr}\{U_p^+ i[U_i,U_k^+]\} = -c_{ki}{}^{p*},\end{aligned} \tag{6.16}$$

whereas (6.9) and (6.11) remain unchanged.

The immediate generalizations of the set of Pauli matrices, $\mathcal{I}$, σ_x, σ_y, and σ_z include: (a) the set of 16

double operators for two-photon polarization $\mathcal{I}^{(A)}\mathcal{I}^{(B)}$, $\mathcal{I}^{(A)}\sigma_x^{(B)}$, $\cdots$, $\sigma_z^{(A)}\sigma_z^{(B)}$, utilized in Sec. 4d, (b) the sets of 16 Dirac operators $\mathcal{I}$, γ_1, $\gamma_2\cdots$, $\gamma_1\gamma_2\gamma_3\gamma_4$ or $\mathcal{I}$, ρ_1, $\rho_1\sigma_x\cdots\rho_3\sigma_z$, (c) the sets of $(2j+1)^2$ multipole operators pertaining to the orientation of particles with spin j, which were referred to in Sec. 4b. The multipole operators are conveniently chosen in tensorial form U_{kq}, i.e. so that they transform under coordinate rotations like the spherical harmonic functions $Y_{kq}(\theta,\varphi)$, with k integer running up to $2j$ and q running in integral steps from $-k$ to k (F53). The $2j+1$ operators with a given k and different q represent the components of the 2^k-pole moment of the particle, to within a normalization constant. Because of the Wigner-Eckart theorem (FR57, Sec. 14), the matrices of the multipole operators are standard Wigner coefficients, properly normalized.

The operators $Q^{(i)}$, defined in Sec. 4g as orthogonal to all powers of the Hamiltonian H, may also conveniently be chosen as orthogonal to one another; together with a set of orthogonal functions of H, they form a complete system. The Wigner phase-space density operators to be discussed in Sec. 9 also constitute a complete orthogonal set.

Orthogonal sets of operators have a role analogous in many respects to the role of orthogonal coordinates in the phase space of classical particle mechanics.[11]

7. HILBERT SPACE REPRESENTATION

Equations (6.4), and (6.2) and (6.3) as well, have the familiar form of scalar products of vectors. A geometrical representation in Hilbert space is indeed appropriate but this space is not the usual Hilbert space of quantum mechanics.

In the usual representation the N eigenstates u_n of (3.1) are taken as base vectors, and all pure states ψ are vectors of unit length of an N-dimensional Hilbert space. The density matrix $\rho_{nn'}$ is a second degree tensor of this space.

Here we consider a Hilbert space with N^2 dimensions, of which the eigenstate products $u_n u_{n'}^+$ constitute one set of base vectors. The density matrix ρ is a *vector* in this space, with the components $\rho_{nn'}$ in the coordinate system of base vectors $u_n u_{n'}^+$. Any operator Q is also represented by a vector, with the components $Q_{n'n}$ in this system, and the mean value $\langle Q\rangle$ may be visualized, according to (3.5), as a scalar product $\mathbf{Q}\cdot\boldsymbol{\varrho}$.

The condition (6.1), with $C=1$, defines the set of operators U_i as an orthonormal system of base vectors, which replaces the system $u_n u_{n'}^+$. The following equations of Sec. 6 simply restate the vector equations of the Hilbert space in terms of the new coordinate systems. In vector notation (6.2), (6.3), and (6.4) read, respectively, $\mathbf{Q}=\sum_i(\mathbf{Q}\cdot\mathbf{U}_i)\mathbf{U}_i=\sum_i Q_i\mathbf{U}_i$, $\boldsymbol{\varrho}=\sum_i(\boldsymbol{\varrho}\cdot\mathbf{U}_i)\mathbf{U}_i=\sum_i\rho_i\mathbf{U}_i$, and $\langle Q\rangle=\boldsymbol{\varrho}\cdot\mathbf{Q}$. All physical equations are real when expressed in terms of the hermitian operators U_i.

Equation (6.6), equivalent to (3.7), fixes the value of the component ρ_0 of $\boldsymbol{\varrho}$, and thus states that the tip of the vector which represents $\boldsymbol{\varrho}$ as drawn from the origin of the Hilbert space lies on the hyperplane of points $\mathbf{x}$ with the coordinate $x_0=[\mathrm{Tr}(\mathcal{I})]^{-\frac{1}{2}}=N^{-\frac{1}{2}}$. The condition (3.11) takes the form $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}\le 1$ and thus limits the tip of $\boldsymbol{\varrho}$ to the interior of a hypersphere of radius 1. The condition $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}=1$ is achieved only for pure states, of which there is only a $(2N-2)$-dimensional manifold, as discussed in Sec. 3g. In mathematical language, this manifold at which $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}$ reaches 1 consists of the "extreme points" of the "convex core" which contains all possible positions of the tip of $\boldsymbol{\varrho}$ (W32, p. 79).

The unitary transformations S of the pure state vectors ψ are represented by all rotations of the ordinary Hilbert space of quantum mechanics. In the Hilbert space considered here the S are also rotations, which transform a vector $\mathbf{Q}$ representing the matrix Q into a vector $\mathbf{Q}'$ representing SQS^{-1}. The S are only a *subgroup* of all rotations in this space, because they are limited by the requirement of leaving invariant the manifold of extreme points $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}=1$. The ordinary Hilbert space is convenient for ordinary quantum mechanical applications because it represents as generic rotations the transformations S, which include the changes of state arising in the course of time from conservative interactions within a system. The representation considered here is convenient for more general applications, in particular for the transformations arising from interactions that are in effect nonconservative (see Secs. 10 and 11).

8. MEASURES OF INFORMATION

Pure states of a system, with $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}=\mathrm{Tr}(\rho^2)=1$ afford a maximum of information; states with $\rho=\rho_0 U_0=[\mathrm{Tr}(\mathcal{I})]^{-1}\mathcal{I}$, and $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}=[\mathrm{Tr}(\mathcal{I})]^{-1}=N^{-1}$, afford a minimum of information. The statistical fluctuations of a physical quantity, represented by $\langle Q^2\rangle-\langle Q\rangle^2=\mathbf{Q}\cdot\boldsymbol{\varrho}-(\mathbf{Q}\cdot\boldsymbol{\varrho})^2$, tend to increase as $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}$ decreases from 1 to N^{-1}, because the second term is quadratic in $\boldsymbol{\varrho}$.

The eigenvalues ρ_j have the distribution $(1,0,0\cdots)$ for pure states and become increasingly uniform as $\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}$ decreases, to the point of being all equal to N^{-1} for states of minimum information. This transition of the spectrum of ρ may be represented by variations of functions of the form

$$\sum_j f(\rho_j)=\mathrm{Tr}[f(\rho)] \tag{8.1}$$

which serve as indices of the quantity of information about the system. A quantity of this type is independent of the representation of ρ and also invariant in the course of time if ρ obeys the Schroedinger equation.

The quantity

$$\sum_j \rho_j^2=\mathrm{Tr}(\rho^2)=\boldsymbol{\varrho}\cdot\boldsymbol{\varrho}\le 1 \tag{8.2}$$

[11] The description of states and of their variations in terms of standard operators and of their mean values can be developed in many directions. For example, Bopp (B56) has been considering standard operators with positive eigenvalues, with the aim of treating the variations of their positive mean values as stochastic processes.

serves itself well as a measure of information, for many purposes. In the representation in terms of orthogonal operators (Sec. 6), we have

$$\mathrm{Tr}(\rho^2)=\sum_{i=0}^{N^2-1}\rho_i^2=N^{-1}+\sum_{i=1}^{N^2-1}\rho_i^2=N^{-1}+\sum_{i=1}^{N^2-1}\langle U_i\rangle^2. \quad (8.3)$$

The quantity of information above the lower limit N^{-1} appears here to consist of additive contributions represented by the squared mean values of operators with trace zero. The spin operators σ of (4.2), the operators $Q^{(v)}$ of (4.12) and $Q^{(i)}$ of (4.18) are all traceless. The degree of polarization P in the spin orientation density matrix (4.2), or in the corresponding light polarization matrix, is closely related to $\mathrm{Tr}(\rho^2)$, since in this example we have

$$\mathrm{Tr}(\rho^2)=\tfrac{1}{2}+\tfrac{1}{2}P^2,\quad P^2=2\sum_i\left(\frac{\sigma_i}{2}\right)^2. \quad (8.4)$$

One may generalize the concept of degree of polarization by defining

$$P^2=\frac{N}{N-1}\sum_{i=1}\rho_i^2. \quad (8.5)$$

Qualitatively, the idea of quantity of information implies that the total information in a number of separate systems is the sum of the quantities of information on the various systems. According to (3.20), the density matrix $\rho^{(ab)}$ of two uncorrelated systems is the product of density matrices $\rho^{(a)}$ and $\rho^{(b)}$ of the two systems. We have then

$$\mathrm{Tr}(\rho^{(ab)2})=\mathrm{Tr}(\rho^{(a)})^2\,\mathrm{Tr}(\rho^{(b)})^2. \quad (8.6)$$

Thus $\ln\mathrm{Tr}(\rho^2)$ is additive and corresponds to the concept of quantity of information better than $\mathrm{Tr}(\rho^2)$ itself. The $\ln\mathrm{Tr}(\rho^2)$ ranges from $-\ln N$ up to 0.

In statistical mechanics it has been known for a long time (N55, T38), that the entropy of a system coincides with the mean value of $-k\ln\rho$,[12] where k is the Boltzmann constant, that is, with

$$\langle -k\ln\rho\rangle=-k\,\mathrm{Tr}(\rho\ln\rho). \quad (8.7)$$

The connection between entropy and quantity of information suggests then that $\mathrm{Tr}(\rho\ln\rho)=\langle\ln\rho\rangle$, or, still better, $\mathrm{Tr}[\rho\ln(N\rho)]=\langle\ln(N\rho)\rangle$, be regarded as a suitable definition for the quantity of information. This definition allows one to introduce a *quantity of information operator* $\ln(N\rho)$, which has itself the additivity property

$$\ln(N_{ab}\rho^{(ab)})=\ln(N_a\rho^{(a)})+\ln(N_b\rho^{(b)}) \quad (8.8)$$

whenever the states of a and b are uncorrelated.

Notice that $\mathrm{Tr}(\rho\ln\rho)$ and $\ln\mathrm{Tr}(\rho^2)$ vary between the same limits and that their values are never very different. Since $\mathrm{Tr}(\rho^2)$ may be expressed as $\langle\rho\rangle$, we are comparing here $\langle\ln\rho\rangle$ with $\ln\langle\rho\rangle$.

The condition $\mathrm{Tr}(\rho\ln\rho)=$ minimum may be utilized in quantum mechanics much like the condition of maximum entropy in thermodynamics or the condition of maximum number of complexions in classical statistical mechanics. The condition $\mathrm{Tr}(\rho\ln\rho)=\min$ with the subsidiary condition that certain operators $L^{(r)}$ have assigned mean values $\langle L^{(r)}\rangle$ determines (N27, S55) that

$$\rho=\exp(\gamma-\textstyle\sum_r\gamma_r L^{(r)}). \quad (8.9)$$

The constants γ, γ_r are fixed as usual by the subsidiary conditions. The state of thermal equilibrium (4.17) is obtained by this procedure, when the set of $L^{(r)}$ consists of the Hamiltonian H only; then $\gamma_1=1/kT$ and $\gamma=-\ln Z$. The discussion of Sec. 4g shows that the condition $\mathrm{Tr}(\rho\ln\rho)=\min$ incorporates into ρ the following information, in addition to the values of the $\langle L^{(r)}\rangle$: (a) all quantities $Q^{(i)}$ such that $\mathrm{Tr}[Q^{(i)}f(L^{(1)}\cdots L^{(r)}\cdots)]=0$ have mean value zero, (b) all quantities $f(L^{(1)}\cdots L^{(r)}\cdots)$ have mean values which are specified functions of the $\langle L^{(r)}\rangle$ and of the parameters γ, γ_r.

9. WIGNER PHASE SPACE DENSITY OPERATORS

Wigner has defined (W32) a quantum analog of the function $\rho(\mathbf{R},\mathbf{P})$ which represents a statistical distribution in phase space of the states of a particle according to classical mechanics. If the state of the particle is identified by the density matrix in the scheme of position coordinates $\mathbf{r}$, $(\mathbf{r}'|\rho|\mathbf{r}'')$, the Wigner distribution function is

$$\rho(\mathbf{R},\mathbf{P})=\frac{1}{(\hbar\pi)^3}\int d\mathbf{r}(\mathbf{R}+\mathbf{r}|\rho|\mathbf{R}-\mathbf{r})\exp(i\hbar^{-1}\mathbf{P}\cdot 2\mathbf{r}). \quad (9.1)$$

This function can be generalized immediately to a many-particle system and has proved useful in statistical mechanics (see, e.g., M48, IZ51).

The values of the function $\rho(\mathbf{R},\mathbf{P})$ constitute a set of parameters $\rho_i=\langle U_i\rangle$, as defined by (6.3). They are the mean values of the operators $U_{\mathbf{RP}}$ with the matrix representation

$$(\mathbf{r}'|U_{\mathbf{RP}}|\mathbf{r}'')=(\hbar\pi)^{-3}\delta(2\mathbf{R}-\mathbf{r}'-\mathbf{r}'')\times\exp[i\hbar^{-1}\mathbf{P}\cdot(\mathbf{r}'-\mathbf{r}'')]. \quad (9.2)$$

($\mathbf{R}$ and $\mathbf{P}$ are parameters, not operators.) The operators $U_{\mathbf{RP}}$ are Hermitian, have trace h^{-3}, and form an orthogonal set, since one verifies from (9.2) that

$$\mathrm{Tr}(U_{\mathbf{R'P'}}U_{\mathbf{RP}})=h^{-3}\delta(\mathbf{R}-\mathbf{R}')\delta(\mathbf{P}-\mathbf{P}'), \quad (9.3)$$

but this set does not include the unit operator. The normalization constant C of (6.1) is, therefore, h^{-3}. Any operator with a matrix $(\mathbf{r}'|Q|\mathbf{r}'')$ can be represented as a linear combination of $U_{\mathbf{RP}}$; the coefficients $h^3\,\mathrm{Tr}(QU_{\mathbf{RP}})$ are calculated by selecting out the matrix elements of Q on the skew diagonal $\mathbf{r}'+\mathbf{r}''=2\mathbf{R}$ and taking their Fourier transform.

[12] The matrix $\ln\rho$ is identified by having the same eigenstates as ρ and the eigenvalues $\ln\rho_j$, equal to the logarithm of the eigenvalues of ρ. Stratonovich (S55) has emphasized that $\mathrm{Tr}(\rho\ln\rho)$ is conveniently expressed as $\sum_j\rho_j\ln\rho_j$ even when ρ is initially identified otherwise than in terms of its eigenvalues and eigenfunctions.

The operator U_{RP} applied to a function $f(\mathbf{r})$ changes it into $(\hbar\pi)^{-3}\exp[i\hbar^{-1}\mathbf{P}\cdot 2(\mathbf{r}-\mathbf{R})]f(2\mathbf{R}-\mathbf{r})$. It follows that U_{RP} has eigenfunctions of the form

$$f(\mathbf{r})=\exp(i\hbar^{-1}\mathbf{P}\cdot\mathbf{r})u^{(\pm)}(\mathbf{r}-\mathbf{R}), \qquad (9.4)$$

where $u^{(\pm)}$ is an even or odd function of its variable. The corresponding eigenvalues are $\pm(\hbar\pi)^{-3}$, respectively. These eigenvalues are highly degenerate; complete systems of orthogonal functions (9.4) can be constructed.[13]

The eigenfunctions (9.4) are wave packets which represent states of the particle with $\langle\mathbf{r}\rangle=\mathbf{R}$ and $\langle\mathbf{p}\rangle=\mathbf{P}$. The existence of negative eigenvalues of U_{RP} constitutes a quantum effect to be contrasted with the condition $\rho(\mathbf{R},\mathbf{P})\geq 0$ obeyed by the classical distribution function.

10. INTERACTION WITH AN UNOBSERVED SYSTEM. IRREVERSIBILITY

When two systems a and b interact, or have been in interaction for a certain period of time, practical interest often centers in the resulting state of a only, irrespective of what has become of b. This state is described at a time t, according to (3.18) and (3.13), by

$$(\rho^{(a)})_t=\mathrm{Tr}_b[(\rho^{(ab)})_t]$$
$$=\mathrm{Tr}_b[\exp(-i\hbar^{-1}Ht)(\rho^{(ab)})_{t=0}\exp(i\hbar^{-1}Ht)], \qquad (10.1)$$

where the Hamiltonian may be regarded as

$$H=H_a+H_b+V, \qquad (10.2)$$

V being the interaction.

The density matrix $\rho^{(a)}$, as defined by (10.1), does not obey the Schroedinger Eq. (3.12) or the equivalent Eq. (6.9). If orthonormal operator sets U_i and V_r are introduced for the systems a and b, respectively, the complete density matrix $\rho^{(ab)}$ is identified by parameters ρ_{ir}, according to (6.3), $\rho^{(ab)}=\sum_{ir}\rho_{ir}U_iV_r$, and $\rho^{(a)}$ by the subset of parameters ρ_{i0},

$$\rho^{(a)}=[\mathrm{Tr}_b(\mathbf{1})]^{\frac{1}{2}}\sum_i\rho_{i0}U_i. \qquad (10.3)$$

The Schroedinger Eq. (6.9) governs the simultaneous variation of all the ρ_{ir}, but we are interested here in the variation of the ρ_{i0} only. In the Hilbert space representation, the complete density matrix is represented by a vector $\boldsymbol{\varrho}^{(ab)}$ which experiences a rotation in the course of time. The components ρ_{i0} define, through (10.3), a vector $\boldsymbol{\varrho}^{(a)}$ which is the projection of $\boldsymbol{\varrho}^{(ab)}$ on the subspace with base vectors U_iV_0. The vector magnitude $\boldsymbol{\varrho}^{(ab)}\cdot\boldsymbol{\varrho}^{(ab)}$ is conserved in the course of time but the magnitude of its projection, $\boldsymbol{\varrho}^{(a)}\cdot\boldsymbol{\varrho}^{(a)}$, which indicates the quantity of information on the system a, is *not conserved*, in general.

In problems of atomic collision often one does not attempt to follow the time variations of $(\rho^{(a)})_t$ but one inquires about the density matrix of the final state f, corresponding to $t=\infty$. The operator $\exp[-i\hbar^{-1}Ht]$ in (10.1) takes then an asymptotic value for $t=\infty$, which may be represented by a scattering or reaction matrix T and (10.1) becomes

$$(\rho^{(a)})_f=\mathrm{Tr}_b(\rho^{(ab)})_f=\mathrm{Tr}_b[T(\rho^{(ab)})_0T^{-1}]. \qquad (10.4)$$

If one calculates $(\rho^{(a)})_t$ or $(\rho^{(a)})_f$ as a matrix, one takes only the trace over eigenstates of b, as indicated in (10.1) and (10.4). Alternately one may want to calculate the corresponding parameters $(\rho_{i0})_t$ or $(\rho_{i0})_f$, or even only the mean value of certain operators $Q^{(a)}$ which are of interest. In this event one takes the trace over eigenstates of both b and a,

$$(\rho_{i0})_t=[\mathrm{Tr}_b(\mathbf{1})]^{-\frac{1}{2}}\mathrm{Tr}_{ab}[U_i\exp(-i\hbar^{-1}Ht)(\rho_{ab})_0 \times\exp(i\hbar Ht)], \qquad (10.5)$$

$$\langle Q^{(a)}\rangle=\mathrm{Tr}_{ab}[Q^{(a)}\exp(-i\hbar^{-1}Ht)(\rho_{ab})_0 \times\exp(i\hbar Ht)]. \qquad (10.6)$$

Equation (10.1) is a suitable point of departure for the description of irreversible processes, such as viscosity or the relaxation of spin orientation in nuclear induction. Quantum mechanics is set up to treat reversible, conservative, processes, and deals in principle with any phenomenon by extending the treatment to a system large enough to be isolated, and hence conservative. Indeed the complete system $a+b$ considered in this section is assumed to be isolated and the variations of $(\rho^{(ab)})_t$ in the course of time are reversible, inasmuch as the initial state description $(\rho^{(ab)})_0$ can be generated mathematically from $(\rho^{(ab)})_t$ by the inverse transformation $\exp(i\hbar^{-1}Ht)(\rho^{(ab)})_t\exp(-i\hbar^{-1}Ht)$.

In the calculation of $(\rho^{(a)})_t$ according to (10.1) much information about the whole system is discarded when one calculates the $\mathrm{Tr}_b[\cdots]$. It is nevertheless implied, in general, that the full matrix $(\rho^{(ab)})_t$, which results from a reversible transformation, has to be calculated prior to the trace calculation. Furthermore, the calculation of $(\rho^{(ab)})_t$ implies, at least in principle, a calculation of $(\rho^{(ab)})_{t'}$ for each t' between 0 and t.

Irreversible processes are special phenomena in which much of the information included in $\rho^{(ab)}$ at a time t' is not relevant to $\rho^{(a)}$ not only at the same instant t' but also at later times $t>t'$. For example, when a nuclear spin $\mathbf{S}_i$, part of a macroscopic amount of matter, is driven to precess by a magnetic field, its precession exerts a magnetic disturbance on the surrounding particles and these disturb, in turn, particles further away. Owing to the very large number of particles, the initial disturbance keeps propagating away and will not react back onto $\mathbf{S}_i$ at later times to any significant extent. The details of this disturbance may accordingly be ignored. That is, one may replace at each time t' the actual $(\rho^{(ab)})_{t'}$ with a simpler $(\rho_s^{(ab)})_{t'}$ which contains less information about b without introducing any significant error in the final result $(\rho^{(a)})_t$ of (10.1). In the Hilbert space representation, the actual vector $\boldsymbol{\varrho}^{(ab)}$ is

[13] Dr. M. S. Green points out that the operator U_{RP} is represented, in terms of projection operators $\mathcal{P}_R^{(\pm)}$ such that $\mathcal{P}_R^{(\pm)}f(r)=(1/2)[f(\mathbf{r})\pm f(2\mathbf{R}-\mathbf{r})]$, by
$U_{RP}=(\hbar\pi)^{-3}\exp(i\hbar^{-1}\mathbf{P}\cdot\mathbf{r})[\mathcal{P}_R^{(+)}-\mathcal{P}_R^{(-)}]\exp(-i\hbar^{-1}\mathbf{P}\cdot\mathbf{r})$.

replaced with a different, "shorter," $\varrho_s^{(ab)}$ which not only has the same projection $\varrho^{(a)}$ but also will evolve in the course of time so that its projection remains effectively the same as that of the correct $\varrho^{(ab)}$. The difference between $\varrho_s^{(ab)}$ and $\varrho^{(ab)}$ is, however, essential from the standpoint of generating back $(\rho^{(ab)})_0$ by inverse transformation; $\exp(i\hbar^{-1}Ht)(\rho_s^{(ab)})_t\exp(-i\hbar^{-1}Ht)$ cannot be expected to resemble $(\rho^{(ab)})_0$. That is to say: the transformation of $(\rho_s^{(ab)})_t$ in the course of time is adequate to represent correctly $(\rho^{(a)})_t=\mathrm{Tr}_b(\rho^{(ab)})_t \sim \mathrm{Tr}_b(\rho_s^{(ab)})_t$ but it is definitely irreversible, whereas the transformation of $(\rho^{(ab)})_t$ is reversible.

In conclusion, the interaction of a with b constitutes an irreversible process when the resulting variation of $\rho^{(a)}$ may be adequately described in terms of an irreversible[14] $(\rho_s^{(ab)})_t$ instead of the correct $(\rho^{(ab)})_t$. Section 11g will indicate a simplification of (10.1) which constitutes an adequate approximation under typical conditions of irreversibility and which leads to a variation of $(\rho^{(a)})_t$ governed by a differential equation. Section 11h will give an equivalent but somewhat more systematic derivation of the same result.

11. APPLICATIONS OF OPERATOR TECHNIQUES

The results of quantum-mechanical calculations can always be expressed by the trace of multiple operator products, like (10.6). Whenever the initial or final states are not specified with a maximum of detail, the final formula of a calculation has in essence the same structure as (10.6), whichever analytical procedure has been employed.

The elementary procedure for calculating a $\mathrm{Tr}(AB\cdots F)$ is to calculate a matrix representation for each factor, A, B, $\cdots F$, and then to work out the multiple sum over products of matrix elements $\sum_{ij\cdots l} A_{ij}B_{jk}\cdots F_{hi}$. This procedure requires one to specify several states with complete sets of quantum numbers, indicated here by i, $j\cdots$, which are usually irrelevant to the problems. Therefore, this procedure not only is laborious but tends to obscure the underlying physical processes and the nature of approximation methods through the introduction of irrelevant variables.

These disadvantages can often be avoided by working on the formal solution of a problem, as indicated in Sec. 10, by a variety of operator techniques. An approach of this kind has been utilized widely in problems involving Dirac spinor variables (see, e.g., H54, p. 216) and is sometimes called the "Casimir trick" (C33, K55). Procedures of operator algebra and calculus (F51) and reduced matrix techniques (FR57) may be utilized in a variety of problems to take advantage of symmetries and invariances and to carry out approximation expansion, with no resort to the explicit matrix representation of operators. The correspondence to the classical variables represented by the operators remains thus in sight.

Many of these procedures have been developed recently and in connection with specialized applications. This section will survey a number of different examples in condensed form, referring to the original literature for details. The last two subsections, relating to irreversible processes with which the author has been particularly concerned, will be developed in somewhat greater detail.

(a) Rotational Invariance. Tensor Operators

For systems effectively isolated in space, the results of a calculation must be independent of the orientation of any coordinates which are utilized. As an example, we consider here the emission of a particle a by a radioactive nucleus which is left in a state b subject to no orientation analysis. We inquire about the angular and spin distribution of the radiation a, on the basis of information on the initial state of the nucleus indicated by ab.

A harmonic analysis of this distribution is performed, in effect, by representing $(\rho^{(a)})_t$ in the form $[\mathrm{Tr}_b\mathit{1}]^{\frac{1}{2}}\sum_i \rho_{i0}U_i^+$, corresponding to (10.3) and (6.14), with the U_i^+ chosen to be *tensor operators*, i.e., such as to transform under coordinate rotations like the spherical harmonic functions $Y_{lm}(\theta,\varphi)$. (Because the Y_{lm} are complex, the transformation law implies that the U_i are not Hermitian, hence the expansion of $(\rho^{(a)})_t$ follows (6.14) instead of (6.3)). The parameters ρ_i vary contragrediently to the U_i^+ and are to be calculated through (10.5). In this equation the Hamiltonian H is invariant under coordinate rotations, if the nucleus is isolated. Since H is invariant and the whole $\mathrm{Tr}_{ab}[\cdots]$ in (10.5) is to be invariant, the initial state matrix $(\rho_{ab})_0$ may then contribute to the trace only through a portion of itself that varies under coordinate rotations contragrediently to U_i, so as to cancel out the variation of U_i. If $(\rho^{(ab)})_0$ is represented, according to (6.14), in the form $\sum_s \rho_s U_s^+$, with the U_s^+ chosen again to vary like spherical harmonics, there will usually be only one operator U_s^+ with $s=s_i$, that transforms contragrediently to U_i, and only this term $s=s_i$ will contribute to $\mathrm{Tr}_{ab}[\cdots]$. (At most there will be a few such terms.) Equation (10.5) reduces then to

$$(\rho_{i0})_t=[\mathrm{Tr}_b(\mathit{1})]^{-\frac{1}{2}}\,\mathrm{Tr}_{ab}[U_i\exp(-i\hbar^{-1}Ht)U_{s_i}\times\exp(i\hbar^{-1}Ht)]\rho_{s_i}. \quad (11.1)$$

Thus each parameter ρ_{i0} of the radiation distribution depends only on the parameter ρ_{s_i} of the initial nuclear state which has the same transformation law under coordinate rotations. This selection rule is qualitatively almost obvious. For example, if the initial nuclear state has axial symmetry—which happens when the nuclear

[14] Van Hove (VH55) has given a quantum mechanical treatment of irreversibility which has elements in common with the developments of Sec. 11g and h but which deals with a somewhat different question. It considers a single large system, such as a and b combined, and inquires about the variations of an operator Q which commutes with an unperturbed Hamiltonian, corresponding to H_a+H_b in (10.2), but not with V. The connections between the Van Hove approach and the approach presented in this review have not been fully explored.

spin is oriented by paramagnetic polarization—all its tensor parameters ρ_s that depend on coordinate rotations about the symmetry axis vanish. Equation (11.1) shows that the corresponding ρ_{i0}'s vanish too. Therefore the radiation distribution has also axial symmetry, which indeed was to be expected.

This analysis of calculations on the basis of rotational transformation properties has been exploited to various extents in applications to the theory of angular correlations and of disintegration processes (TG49, F51a, D52, WA52, CJ53, F53, W53, W54b), and is developed systematically by tensor algebra in (FR57, particularly Sec. 20).

(b) Depolarization Coefficients

An analysis based on tensor algebra also exploits the following consideration, for the calculation of $\mathrm{Tr}_{ab}[\cdots]$ in (11.1). Since the angular momentum is a constant of the motion, the Hamiltonian operator is diagonal in the scheme of eigenstates with quantum number j, where j is the spin of the initial nuclear state. It is then convenient to represent all operators in this scheme. The operators U_i of the emitted radiation are initially expressed in the scheme of angular momentum quantum numbers j_r of the radiation. The transformation from the scheme j_r to the scheme j depends on the spin j' of the residual nucleus, owing to the conservation of angular momentum in the disintegration, $\mathbf{J}=\mathbf{J}'+\mathbf{J}_r$, but involves only geometrical relations of coupling independent of the physical process of disintegration. When no observation is performed on the orientation of the residual spin, the transformation of U_i to the scheme j consists in essence of multiplying its matrix by a Racah coefficient $\overline{W}$ which is a function of the quantum numbers j, j_r, and j' and of the degree of the spherical harmonic with transforms like U_i. The factor $\overline{W}$ thus introduced in (11.1) is no larger than 1 in absolute value and thus reduces the value of the parameter $(\rho_{i0})_t$ which is a coefficient of the harmonic analysis of the radiation distribution. The larger the spin j' of the residual nucleus the smaller are $\overline{W}$ and $(\rho_{i0})_t$, so that the radiation distribution is more smeared out. This was also to be expected because more potential information has been lost by failing to observe the orientation of this nucleus. Therefore $\overline{W}$ may be called a *depolarization coefficient*. It indicates how much of the information originally available on the orientation of the nucleus fails to express itself in the radiation distribution owing to failure to observe the residual nucleus.

The experiment on light emission discussed in Sec. 5 is, of course, quite analogous to the nuclear disintegration process considered here. Equation (5.1), expressed in terms of spherical harmonics, has the form

$$I(\theta)\rho\propto\frac{10}{9}\left\{\left[1-\tfrac{1}{10}P_{20}(\theta)\right]\mathbb{1}+\frac{6^{\frac12}}{10}P_{22}(\theta)\sigma_\zeta\right\},\quad(11.2)$$

with $P_{20}=(3/2)\cos^2\theta-(\frac12)$ and $P_{22}=(3/8)^{\frac12}\sin^2\theta$. In the calculation of (11.2) the coefficients of P_{20} and P_{22} have been reduced by a factor $\frac{1}{10}$ owing to a $\overline{W}$ coefficient which takes into account the quantum number $j'=2$ of the atom after the radiation emission. If the atom were left in an s state ($j'=0$), this factor $\frac{1}{10}$ would not appear; indeed the light emitted in a transition from ($l=1$, $m=0$) to $l=0$ is fully polarized and its angular distribution is well modulated, as indicated by

$$I(\theta)\rho\propto\left[1-P_{20}(\theta)\right]\mathbb{1}+6^{\frac12}P_{22}(\theta)\sigma_\zeta=\tfrac{3}{2}\sin^2\theta(\mathbb{1}+\sigma_\zeta).\quad(11.3)$$

(c) Interaction Representation—Feynman Calculus

Equation (3.18) shows that the calculations of a mean value $\langle Q\rangle$ in the Heisenberg representation ($Q=Q(t)$, $\rho=$const) and in the Schroedinger representation ($Q=$const, $\rho=\rho(t)$) are manifestly equivalent. When the Hamiltonian consists of two or more terms, as in (10.2), other representations are possible, in which the time variations induced by different terms of the Hamiltonian are separated out. The Hamiltonian (10.2) contains independent-system terms H_a+H_b and a coupling term V. In the "interaction representation" the variations induced by H_a+H_b may be applied to Q and the residual influence of V applied to ρ. This representation is particularly convenient when $Q=Q^{(a)}$ and commutes with H_b, because H_b causes then no variation of Q.

The formal derivation and the applications of the interaction representation are facilitated by Feynman's operator calculus (F51) which shows how to "disentangle" operators contained in the same exponential function, to find

$$\exp[-i\hbar^{-1}(H_a+H_b+V)t]$$
$$=\exp[-i\hbar^{-1}(H_a+H_b)t]$$
$$\times\exp\left[-i\hbar^{-1}\int_0^t V(t')dt'\right],\quad(11.4)$$

with

$$V(t)=\exp[i\hbar^{-1}(H_a+H_b)t]V$$
$$\times\exp[-i\hbar^{-1}(H_a+H_b)t],\quad(11.5)$$

and the reciprocal formula

$$\exp[i\hbar^{-1}(H_a+H_b+V)t]$$
$$=\exp\left[i\hbar^{-1}\int_0^t V(t')dt'\right]\exp[i\hbar^{-1}(H_a+H_b)t].\quad(11.6)$$

Equation (10.6) becomes, in this representation,

$$\langle Q^{(a)}\rangle=\mathrm{Tr}_{ab}\left\{Q^{(a)}(t)\exp\left[-i\hbar^{-1}\int_0^t V(t')dt'\right](\rho^{(ab)})_0\right.$$
$$\left.\times\exp\left[i\hbar^{-1}\int_0^t V(t')dt'\right]\right\},\quad(11.7)$$

where $Q^{(a)}(t)$ is independent of H_b, and (10.1) becomes

$$(\rho^{(a)})_t=\exp(-i\hbar^{-1}H_a t)$$
$$\times\mathrm{Tr}_b\left\{\exp\left[-i\hbar^{-1}\int_0^t V(t')dt'\right](\rho^{(ab)})_0\right.$$
$$\left.\times\exp\left[i\hbar\int_0^t V(t')dt'\right]\right\}\exp(i\hbar^{-1}H_a t), \quad (11.8)$$

where H_b has dropped out from the expression of the trace.

Feynman (F51) has also pointed out how, given an operator, such as V, acting on both systems a and b, one may calculate its matrix elements between states of b only, so that these matrix elements are still operators with regard to a. To this end one may expand the operator V into operators U_i of a, according to (6.2), $V=\sum_i V_i U_i$, the coefficients V_i now being operators of b only. We have then

$$V(t)=\sum_i V_i(t)U_i(t)=\sum_i \exp(i\hbar^{-1}H_b t)V_i$$
$$\times\exp(-i\hbar^{-1}H_b t)\exp(i\hbar^{-1}H_a t)U_i$$
$$\times\exp(-i\hbar^{-1}H_a t), \quad (11.9)$$

which enables us to calculate directly the matrix elements of $V_i(t)$ only, or also expressions such as the $\mathrm{Tr}_b[\cdots]$ in (11.8).

In the calculation pertaining to b one may treat the operators $U_i(t)$ as though they were ordinary functions rather than operators. A main point of Feynman's method lies in utilizing variable indices, such as t, to keep track of the correct ordering of the operators of the system a, even though the usual operator algebra rules may have been formally disregarded in the calculation relating to b.

(d) Perturbation Expansions

The results of the usual time-dependent perturbation theory are easily obtained in a more general form by utilizing the expansion (F51)

$$\exp\left[\int_0^t A(t')dt'\right]=1+\int_0^t A(t')dt'$$
$$+\int_0^t dt'\int_0^{t'} dt'' A(t')A(t'')+\cdots, \quad (11.10)$$

which enables one to substitute in (11.7) or (11.8),

$$\exp\left[-i\hbar^{-1}\int_0^t V(t')dt'\right](\rho^{(ab)})_0\exp\left[i\hbar^{-1}\int_0^t V(t')dt'\right]$$
$$=(\rho^{(ab)})_0-i\hbar^{-1}\int_0^t dt'[V(t'),(\rho^{(ab)})_0]$$
$$-\hbar^{-2}\int_0^t dt'\int_0^{t'} dt''[V(t'),[V(t''),(\rho^{(ab)})_0]]+\cdots. \quad (11.11)$$

The usual perturbation formulas are obtained by representing (11.11) in the scheme of eigenstates of H_A+H_B; the matrix elements of $V(t)$ are then simple exponentials in t, as in (3.15), and the integrations over t' and t'' may be carried out analytically.

(e) Correlation Functions

When the systems a and b are initially independent, i.e. when $(\rho^{(ab)})_0=(\rho^{(a)})_0(\rho^{(b)})_0$, V is expanded in the form (11.9), and the perturbation expansion (11.11) is utilized, the $\mathrm{Tr}_b[\cdots]$ in (11.7) or (11.8) depends on b through a series of terms of the form

$$\mathrm{Tr}\{V_i(t')V_j(t'')\cdots V_l(t^{(n)})(\rho^{(b)})_0\}$$
$$=\langle V_i(t')V_j(t'')\cdots V_l(t^{(n)})\rangle. \quad (11.12)$$

These parameters are time-correlation functions of a kind often considered in statistical mechanics, that is, mean values of products of physical quantities taken at different instants of time. (However, in classical mechanics the averaging is usually done by an integration over time intervals, whereas in (11.12) we deal with an ensemble average.)

In the frequent case where b is initially in a steady state, i.e., when $(\rho^{(b)})_0$ commutes with H_b, the correlation functions (11.12) depend only on the time intervals $t'-t''$, $t''-t'''$, etc. but not on the absolute values of t', $t''\cdots$. In this event $\langle V_i(t')\rangle$ is altogether independent of t', and the pair correlation function $\langle V_i(t')V_j(t'')\rangle$ depends only on $t'-t''$ and may be written as $\langle V_iV_j(t''-t')\rangle$.

If the expansion $V=\sum_i U_iV_i$ is not utilized, we have instead of (11.12) expressions of the form $\mathrm{Tr}_b\{V(t')V(t'')\cdots V(t^{(n)})(\rho^{(b)})_0\}$ which are operators of the system a, whereas the parameters (11.12) are numerical coefficients of the operators $U_i(t')U_j(t'')\cdots$. If $(\rho^{(b)})_0$ commutes with H_b, the dependence on the absolute time scale can still be brought out of the correlation function; we have then

$$\mathrm{Tr}\{V(t')V(t'')\cdots V(t^{(n)})(\rho^{(b)})_0\}$$
$$=\exp(i\hbar^{-1}H_a t')\,\mathrm{Tr}_b\{VV(t''-t')\cdots V(t^{(n)}-t')(\rho^{(b)})_0\}$$
$$\cdot\exp(-i\hbar^{-1}H_a t'). \quad (11.13)$$

The correlation functions are proving increasingly useful for representing and classifying the properties of a system which are relevant to its interaction with other systems. For a system initially in a steady state, the static parameters $\langle V_i\rangle$ represent constant fields that produce elastic scattering effects, whereas the pair correlations $\langle V_i(0)V_j(t)\rangle$ represent to first-order inelastic interaction parameters. For example, Van Hove (VH54) has represented the neutron scattering properties of a crystal in terms of a pair correlation between the positions of nuclei $\mathbf{r}_i$, $\mathbf{r}_j$ in the crystal

$$G(\mathbf{r},t)=N^{-1}\left(\sum_{i,j=1}^{N}\int d\mathbf{r}'\delta(\mathbf{r}+\mathbf{r}_i(0)-\mathbf{r}')\right.$$
$$\left.\times\delta(\mathbf{r}'-\mathbf{r}_j(t))\right). \quad (11.14)$$

The function $G(\mathbf{r},t)$ is a macroscopic property of the crystal, but its definition is quantum-mechanical, atomistic, rather than phenomenologic. As a further example, the dielectric constant of a material, determined by electron polarizability, is defined, in terms of the $G(\mathbf{r},t)$ function for electrons (F56), by

$$\epsilon(\omega)=\left\{1+\frac{Ne^2}{\hbar}\int_0^\infty dt e^{-i\omega t}\int d\mathbf{r} \times\int d\mathbf{r}'\frac{G(\mathbf{r}',t)-G(\mathbf{r}',t)^*}{|\mathbf{r}-\mathbf{r}'|}\right\}^{-1}. \quad (11.15)$$

(f) Probability of Energy or Momentum Transfers

There is another approach which leads one to express the probability of an interaction effect in terms of a correlation function (AP47). Suppose one inquires about the probability $P(\omega)d\omega$ that system a transfers to b an energy amount between $\hbar\omega$ and $\hbar(\omega+d\omega)$ in the time interval from 0 to t, irrespective of other effects of the interaction. This probability is represented by the mean value of an operator $P(\omega)$ which is diagonal in the scheme of energy eigenstates of a and has in this scheme the form $\delta(E_{at}-E_{a0}+\hbar\omega)$, where E_{at} and E_{a0} are eigenvalues of H_a pertaining, respectively, to states at the times t and 0. The mean value of P is to be evaluated according to (10.6). We represent $P(\omega)$ in the form

$$P(\omega)=\delta(E_{at}-E_{a0}+\hbar\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau \exp(i\omega\tau) \times\exp(i\hbar^{-1}E_{at}\tau)\exp(-i\hbar^{-1}E_{a0}\tau) \quad (11.16)$$

and notice that both E_{at} and E_{a0} may be represented by the Hamiltonian H_a° operating, respectively, after and before the interaction. Entering into (10.6) $Q^{(a)}=P(\omega)$, we may write, in operator notation independent of a specific matrix representation,

$$\langle P(\omega)\rangle=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau \exp(i\omega\tau) \times\mathrm{Tr}_{ab}\{\exp(i\hbar^{-1}H_a\tau)\exp(-i\hbar^{-1}Ht) \times\exp(-i\hbar^{-1}H_a\tau)(\rho^{(ab)})\exp(i\hbar^{-1}Ht)\}. \quad (11.17)$$

Further, utilizing the interaction representation (11.7) and the expansion (11.9), and considering that for any operator $F^{(a)}(t)$ in the interaction representation $\exp(i\hbar^{-1}H_a\tau)F^{(a)}(t)\exp(-i\hbar^{-1}H_a\tau)=F^{(a)}(t+\tau)$, we have

$$\langle P(\omega)\rangle=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau \exp(i\omega\tau) \times\mathrm{Tr}\left\{\exp\left[-i\hbar^{-1}\int_0^t\sum_i U_i(t'+\tau)V_i(t')\right](\rho^{(ab)})_0 \times\exp\left[i\hbar^{-1}\int_0^t\sum_j U_j(t')V_j(t')\right]\right\}. \quad (11.18)$$

In the first Born approximation the integrations over t' are carried out through the second order term of the perturbation expansion (11.11), and (11.18) becomes

$$\langle P^{(a)}(\omega)\rangle=\frac{t}{\hbar}\int_{-\infty}^{\infty}d\tau \exp(i\omega\tau)\,\mathrm{Tr}_{ab}\{[\sum_i \bar{U}_i(\tau)\bar{V}_i] \times(\rho^{(ab)})_0[\sum_j \bar{U}_j\bar{V}_j]\}, \quad (11.19)$$

where $\bar{U}_i\bar{V}_i$ represents that portion of the operator product U_iV_i which is diagonal in the total energy and induces transitions with energy conservation, namely

$$\bar{U}_i\bar{V}_i=(2\pi\hbar)^{-1}\int_{-\infty}^{\infty}dt U_i(t)V_i(t). \quad (11.20)$$

It follows that $\bar{U}_i(\tau)\bar{V}_i=\bar{U}_i\bar{V}_i(-\tau)$.

Expressions like (11.19) and (11.20) show explicitly how transition probabilities depend on a time variable whose role is obscured in the usual quantum-mechanical formulas. For example, the absorption or emission of light by an atomic electron depends on the correlation between the electron's position $\mathbf{r}$ or velocity $\mathbf{v}$ at time intervals τ. (In this problem, we have $\sum_i U_iV_i=e\mathbf{v}\cdot\mathbf{A}(\mathbf{r})$, where $\mathbf{A}$ is the vector potential.) The probability of emitting or absorbing a photon $\hbar\omega$ is proportional to the spectral component ω of the correlation function. Van Hove was led to consider the correlation function G of (11.14) (VH54) through a formula similar to (11.19). Wick (W54a) utilized such a formula as the point of departure of an approximate calculation of neutron scattering in which the correlation function is expanded into powers of τ, to reflect the fact that a neutron scattering process is short as compared to the vibration period of a crystal lattice.

The probability of a momentum transfer $\hbar\mathbf{q}$ may be treated much like that of an energy transfer, in terms of an operator $\delta(\mathbf{p}_{at}-\mathbf{p}_{a0}+\hbar\mathbf{q})$ expanded into a triple Fourier integral. The operators $\exp(\pm i\hbar^{-1}\mathbf{p}_a\cdot\boldsymbol{\xi})$, which replace $\exp(\pm i\hbar^{-1}H_a\tau)$ in (11.17) have the effect of shifting all space coordinates in the operator U_i by $\boldsymbol{\xi}$, so that the trace takes the form of a space correlation function of the type $\langle f(\mathbf{r})f(\mathbf{r}+\boldsymbol{\xi})\rangle$.

(g) Wangsness-Bloch Equation for Irreversible Processes

Wangsness and Bloch (WB53) treated the relaxation of nuclear spin orientation, in essence, along the lines indicated at the end of Sec. 10. The variable orientation of the spin of a fixed atomic nucleus constitutes the system a, whose Hamiltonian H_a represents the coupling with an external magnetic field $\mathbf{H}$, and the system of all surrounding particles constitutes b, the Hamiltonian H_b being unspecified. At the initial time $t=0$, the states of the two systems are assumed to be uncorrelated, so that $(\rho^{(ab)})_0=(\rho^{(a)})_0(\rho^{(b)})_0$, according to (3.20), with $(\rho^{(a)})_0$ unspecified and $(\rho^{(b)})_0$ assumed to be a state of thermal equilibrium according to (4.17). The coupling V of a and

b is also left unspecified except that $\mathrm{Tr}_b V$ is assumed to vanish, though only for simplicity.

The state of spin orientation at a somewhat later time $t=\tau$, represented by $(\rho^{(a)})_\tau$, is then given by (11.8) but τ is assumed to be sufficiently small to warrant application of the perturbation expansion (11.11) extended to second order terms only. Disregarding further some minor side effects, one finds

$$(\rho^{(a)})_\tau = \exp(-i\hbar^{-1}H_a\tau)\,\mathrm{Tr}_b\left\{(\rho^{(a)})_0(\rho^{(b)})_0 - \frac{\pi\tau}{\hbar}[\bar{V},[\bar{V},(\rho^{(a)})_0(\rho^{(b)})_0]]\right\}\exp(i\hbar^{-1}H_a\tau), \quad (11.21)$$

where $\bar{V}$ represents that portion of the operator V which is diagonal in the total energy and is defined according to (11.20). (The first-order term is missing because $\mathrm{Tr}_b V$ is assumed to vanish.)

To extend the calculation to larger intervals $t>\tau$ one would need, in general, to know not only $(\rho^{(a)})_\tau$ but the density matrix of the whole system $(\rho^{(ab)})_\tau$. At this point Wangsness and Bloch take into account that b is a large system through which the effects of interaction with the spin a dissipate away quickly. Therefore, it is argued, the state of b will not have changed significantly during τ. The actual density matrix $(\rho^{(ab)})_\tau$, resulting from $(\rho^{(a)})_0(\rho^{(b)})_0$ by reversible transformation, is replaced with

$$(\rho_s^{(ab)})_\tau = (\rho^{(a)})_\tau(\rho^{(b)})_0. \quad (11.22)$$

The evolution in time of the combined system becomes thereby irreversible.

If this $(\rho_s^{(ab)})_\tau$ is entered into (11.21) in the place of $(\rho^{(a)})_0(\rho^{(b)})_0$, the result is $(\rho^{(a)})_{2\tau}$. The operation can be repeated chainwise yielding $(\rho^{(a)})_t$ at successive time intervals $\tau, 2\tau, 3\tau, \cdots n\tau\cdots$. This repetitive operation is analogous to the numerical integration of a differential equation. Indeed, the variations of $(\rho^{(a)})_t$ thus obtained obey a differential equation, where the interval τ is treated formally as a differential dt[15]

$$\frac{d\rho^{(a)}}{dt} = -\frac{i}{\hbar}[H_a,\rho^{(a)}] - \frac{\pi}{\hbar}\mathrm{Tr}_b[\bar{V},[\bar{V},\rho^{(a)}(\rho^{(b)})_0]]. \quad (11.23)$$

Wangsness and Bloch (WB) pointed out that (11.23) is a quantum analog of the Boltzmann equation of statistical mechanics, and generalizes the Schroedinger equation to incorporate irreversible transformations. Expansion of $\rho^{(a)}$ into orthogonal operators, $\rho^{(a)} = \sum_i \rho_i U_i$, brings the WB equation to a form analogous to (6.9),

$$d\rho_i/dt = \sum_k(\Omega_{ik}+M_{ik})\rho_k, \quad (11.24)$$

with Ω_{ik} as in (6.11) and

$$M_{ik} = -\frac{\pi}{\hbar}\mathrm{Tr}_{ab}\{U_i[\bar{V},[\bar{V},U_k(\rho^{(b)})_0]]\}. \quad (11.25)$$

Equation (11.24) represents a transformation of the vector $\boldsymbol{\varrho}^{(a)}$ in the Hilbert space of Sec. 7, but we deal no longer with a simple rotation because M_{ik} is not antisymmetric. With $U_i = U_0 \propto 1$, we have $M_{0k}=\Omega_{0k}=0$ (but $M_{i0}\neq 0$ in general), so that $\partial\rho_0/\partial t=0$ as required; it also follows that the determinant of $\Omega_{ik}+M_{ik}$ vanishes, so that (11.24) has a steady state solution.

Further expansion $\bar{V}=\sum_r U_r\bar{V}_r$ makes it possible to represent the matrix M_{ik} in terms of separate properties of the systems a and b. The relevant properties of a depend only on its mechanical structure as represented by relationships among the operators U_i. They depend in particular on the coefficients $c_{ik}{}^p$ of (6.10) and on the corresponding symmetric coefficients.

$$s_{ik}{}^p = \mathrm{Tr}[(U_iU_k+U_kU_i)U_p]. \quad (11.26)$$

Entering the expansion of $\bar{V}$ into (11.25), working out the double commutator, and separating out symmetric and antisymmetric terms, one finds

$$M_{ik} = -\sum_{rs}\left\{A_{ikrs}\frac{\pi}{2\hbar}\langle\bar{V}_r\bar{V}_s+\bar{V}_s\bar{V}_r\rangle + B_{ikrs}\frac{\pi}{2\hbar}\langle i[\bar{V}_r,\bar{V}_s]\rangle\right\} \quad (11.27)$$

where

$$A_{ikrs} = A_{kisr} = \sum_p c_{ir}{}^p c_{ks}{}^p, \quad (11.28)$$

$$B_{ikrs} = -B_{iksr} = \sum_p c_{ir}{}^p s_{ks}{}^p = \sum_p(\tfrac{1}{2}s_{ik}{}^p c_{rs}{}^p + \tfrac{1}{4}c_{ir}{}^p s_{ks}{}^p - \tfrac{1}{4}c_{kr}{}^p s_{is}{}^p - \tfrac{1}{4}c_{is}{}^p s_{kr}{}^p + \tfrac{1}{4}c_{ks}{}^p s_{ir}{}^p). \quad (11.29)$$

The coefficients A and B represent the properties of a. They obey selection rules, arising from the invariance of traces under coordinate rotation or under reflection in space or time. For example, the $c_{ik}{}^p$ vanish unless the product of the parities of U_i, U_k and U_p is odd, and the $s_{ik}{}^p$ vanish unless it is even, so that $A_{ikrs}=0$ if the product $U_iU_kU_rU_s$ is odd and $B_{ikrs}=0$ when it is even.

The mean values $\langle\bar{V}_r\bar{V}_s+\bar{V}_s\bar{V}_r\rangle$ and $\langle i[\bar{V}_r,\bar{V}_s]\rangle$ represent the relevant properties of b. They are essentially equivalent to the admittance matrix, with rows and columns indicated by r and s, which has been introduced by Callen and co-workers (CW51, CBJ52). These authors considered the reaction of a large system, like b, to an external time variable potential $Q_s\exp(i\omega t)$, representing the reaction by the mean value of the rate of change of another operator $\dot{Q}_r$, as calculated by first-order perturbation. The Feynman calculus exploits the analogy between the reaction of b to an external po-

[15] In the derivation of (11.21) it is assumed that τ is not infinitesimal but, rather, that is sufficiently large for the spectrum of the interaction, $\int_{-\infty}^{\infty}V(t)\exp(-i\omega t)dt$ to be effectively constant over an interval $\delta\omega=\tau^{-1}$. Therefore (11.21) and (11.23) represent correctly only semimacroscopic time variations of $\rho^{(a)}$, leaving out variations on a finer scale of time intervals $\lesssim(\delta\omega)^{-1}$.

tential and to the interaction V with another quantum-mechanical system a, by permitting one to calculate the reaction to a without considering at every step that V is also an operator of a.[16]

A system a with N orthogonal eigenstates has N^2 operators U_i, so that the matrix $\Omega_{ik}+M_{ik}$ has N^2 rows and columns. The relevant characteristics of b are represented by N^4 coefficients $\langle \bar{V}_r\bar{V}_s+\bar{V}_s\bar{V}_r\rangle$ and as many $\langle i[\bar{V}_r,\bar{V}_s]\rangle$. The matrix $\Omega_{ik}+M_{ik}$ has N^2 complex eigenvalues, whose real parts are time-relaxation constants; one of these eigenvalues vanishes, in correspondence to the steady states, the remaining N^2-1 relaxation constants determine the rate of approach to the steady state. All these parameters need not be different and their number is often greatly reduced owing to symmetry.

In the nuclear induction with spin $\frac{1}{2}$, there could be $N^2-1=3$ relaxation constants, but axial symmetry about the main magnetic field reduces this number to 2, corresponding to the well-known longitudinal and transverse relaxation times. For spin 1 we have $N^2-1=8$ constants, reduced to 5 by axial symmetry. However Wangsness and Bloch argue that dipole and quadrupole polarizations are effectively uncoupled in this case.[17] Hence the magnetic dipole polarization, which is responsible for nuclear induction, behaves for $j=1$ as though j were $\frac{1}{2}$.

(h) Irreversibility Approximation by Expansion in Cumulants

Instead of introducing irreversibility through the convention (11.22), one may formulate a systematic approximation procedure (F54b), which, when applied to first order, reduces the exact expression (11.8) for $(\rho^a)_t$ to a solution of the Wangsness-Bloch Eq. (11.23). Qualitatively, if a system b is large and quickly dissipates the effect of disturbances, one expects b to show a "lack of memory" in its interactions with other systems, such that the time variations of an operator $V(t)$ in the neighborhood of time instants t' and t'' would be uncorrelated, for $t'-t''$ sufficiently large. Mathematically, this lack of correlation should cause the mean value of an operator such as $\langle V_i(t')V_j(t'')\cdots V_l(t^n)\rangle$ in (11.12) to approach rapidly the product of mean values $\langle V_i\rangle\langle V_j\rangle\langle V_l\rangle$ as the time intervals $t'-t''\cdots t'-t^{(n)}$ increase.

In order to take advantage of this circumstance one considers in particular the departures of an operator from its mean value, defining for example $\mathcal{V}(t)=V(t)-\langle V(t)\rangle_b$ (where $\langle A\rangle_b$ means Tr_bA), so that $\langle\mathcal{V}(t)\rangle_b=0$

$$V(t)=\langle V(t)\rangle_b+\mathcal{V}(t) \tag{11.30}$$

and

$$\langle VV(t)\rangle_b=\langle V\rangle_b\langle V(t)\rangle_b+\langle\mathcal{V}\mathcal{V}(t)\rangle_b. \tag{11.31}$$

The mean value $\langle\mathcal{V}\mathcal{V}(t)\rangle_b$, is the correlation of "deviates." It represents the correlation effect proper and is expected to vanish when t increases above an effective memory range, or relaxation time, of b. This procedure may be continued by separating out deviates of successively higher orders, defining, for example,

$$\mathcal{V}(t')\mathcal{V}(t'')=\langle\mathcal{V}(t')\mathcal{V}(t'')\rangle_b+\mathcal{W}(t')\mathcal{W}(t''), \tag{11.32}$$

so that the deviates $\mathcal{W}$ have not only $\langle\mathcal{W}(t)\rangle_b=0$ but also $\langle\mathcal{W}(t')\mathcal{W}(t'')\rangle_b=0$; the $\mathcal{W}$'s contribute only to quadruple and higher correlations.

When expressions like (11.30) and (11.32) are entered in the perturbation expansion (11.11), the mean values of different orders $\langle V(t)\rangle_b$, $\langle\mathcal{V}(t')\mathcal{V}(t'')\rangle_b$, etc. can be sorted out from the expansion. Leaving out the complications due to commutation with $(\rho^{(ab)})_0$ (see F54b), and using (11.13) the reader may verify that the simpler expansion

$$\mathrm{Tr}\left\{\exp\left(-i\hbar^{-1}\int_0^t V(t')dt'\right)\rho^{(b)}\right\}$$
$$=\mathrm{Tr}_b\left\{\left[1-i\hbar^{-1}\int_0^t V(t')dt'\right.\right.$$
$$\left.\left.-\hbar^{-2}\int_0^t dt'\int_0^{t'} dt''V(t')V(t'')+\cdots\right]\rho^{(b)}\right\} \tag{11.33}$$

condenses to the exponential form

$$\exp\left\{-i\hbar^{-1}\int_0^t dt'\exp(i\hbar^{-1}H_at')\right.$$
$$\times\left[\langle V\rangle_b-i\hbar^{-1}\int_0^{t'}dt''\langle\mathcal{V}\mathcal{V}(t''-t')\rangle_b\right.$$
$$\left.-\hbar^{-2}\int_0^{t'}dt''\int_0^{t''}dt'''\langle\mathcal{V}\mathcal{V}(t''-t')\mathcal{V}(t'''-t')\rangle_b+\cdots\right]$$
$$\left.\times\exp(-i\hbar^{-1}H_at')\right\}. \tag{11.34}$$

The average of the exponential operator in (11.33) has been thereby reduced to the exponential function of a sequence of averaged operators, which are correlation functions of increasing order. The reduction of (11.33) to (11.34) is analogous to the equation of probability theory $\langle\exp(t\xi)\rangle=\exp(\sum_\nu t^\nu\kappa_\nu/\nu!)$, where ξ is a random variable and the κ_ν are the cumulants $\langle\xi\rangle$, $\langle(\xi-\langle\xi\rangle)^2\rangle$, $\langle(\xi-\langle\xi\rangle)^3\rangle\cdots$ of its distribution (C46).

[16] The analog of Callen's time variation $\exp(i\omega t)$ of the external disturbance is included in the WB problem through the definition

$$V_s(t)=\mathrm{Tr}_a\{U_sV(t)\}=\mathrm{Tr}\{\exp(i\hbar^{-1}H_at)U_s\exp(-i\hbar^{-1}H_at)\times\exp(i\hbar^{-1}H_bt)V\exp(-i\hbar^{-1}H_bt)\}$$

which contains a factor depending on all eigenfrequencies of H_a.

[17] The coefficients $\langle\bar{V}_r\bar{V}_s+\bar{V}_s\bar{V}_r\rangle$ and $\langle i[\bar{V}_r,\bar{V}_s]\rangle$ can be classified according to the multipolarity of the operators U_r and U_s. If U_r is a dipole and U_s a quadrupole operator, $\bar{V}_r$ represents a magnetic field and $\bar{V}_s$ an electric field, and WB assume that the fluctuations of these different fields are uncorrelated so that the mixed-field coefficients vanish. It follows then from the parity selection rules of the coefficients A_{ikrs} and B_{ikrs} that the parameters ρ_i and ρ_k pertaining to magnetic and electric polarizations will remain uncoupled.

The expansion (11.34) affords a suitable point of departure for an irreversibility approximation. Three steps are involved in deriving from (11.34) a simpler expression equivalent to the Wangsness-Bloch Eq. (11.23): (1) It must be assumed that, because of the internal mechanics of system b, the correlation functions in (11.34) vanish rapidly as $t'-t''$, $t'-t'''\cdots$ increase, so that, as t' increases, the integrals in the brackets reach rapidly a limiting value corresponding to $t'=\infty$ and thus are no longer explicit functions of t' (short-memory approximation). (2) The whole expression in the brackets can be handled as a single operator of the system a, acting at the time t', even though some portions of it actually should operate at times t'', $t'''\cdots$ somewhat earlier than t'. Disregarding these fine distinctions in the time scale is equivalent to treating the time intervals τ as differentials in the derivation of (11.23). (3) The terms of third and higher order can be disregarded. This assumption relates to the short-memory approximation (1) because high-order interaction effects would build up if allowed to accumulate for a long time. It implies that second-order perturbation treatment is adequate, as in (11.21), to describe the effects of interaction upon a during each memory interval.

Thus the whole effect of irreversible interaction with system b is represented by pair correlation integrals of the type $\int_0^\infty dt\langle \mathcal{V}\mathcal{V}(-t)\rangle_b$. Upon expansion into orthogonal operators $\mathcal{V}(t)=\sum_s U_s\mathcal{V}_s(t)$ these integrals depend on b through mean values $\langle \mathcal{V}_r\mathcal{V}_s(-t)\rangle$ which are again equivalent to Callen's admittance matrix mentioned at the end of Sec. 11g.

REFERENCES

AP47 A. Akhiezer and I. Pomeranchuk, J. Phys. (U.S.S.R.) **11**, 167 (1947).
B46 F. Bloch, Phys. Rev. **70**, 460 (1946).
B56 F. Bopp, Z. Physik **144**, 13 (1956).
BR45 F. Bloch and I. I. Rabi, Revs. Modern Phys. **17**, 237 (1945).
C33 H. B. G. Casimir, Helv. Phys. Acta **6**, 287 (1933).
C46 H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946), pp. 185–186.
CBJ52 Callen, Barasch, and Jackson, Phys. Rev. **88**, 1382 (1952).
CJ53 F. Coester and J. Jauch, Helv. Phys. Acta **26**, 3 (1953).
CW51 H. B. Callen and T. A. Welton, Phys. Rev. **83**, 34 (1951).
D52 R. H. Dalitz, Proc. Phys. Soc. (London) **A65**, 175 (1952).
E37 W. M. Elsasser, Phys. Rev. **52**, 987 (1937).
F49 U. Fano, J. Opt. Soc. Am. **39**, 859 (1949).
F51 R. Feynman, Phys. Rev. **84**, 108 (1951).
F51a U. Fano, NBS Report 1214, unpublished (1951).
F53 U. Fano, Phys. Rev. **90**, 577 (1953).
F54a U. Fano, Phys. Rev. **93**, 121 (1954).
F54b U. Fano, Phys. Rev. **96**, 869 (1954).
F56 U. Fano, Phys. Rev. **103**, 1202 (1956).
FMcD51 D. L. Falkoff and J. E. McDonald, J. Opt. Soc. Am. **41**, 862 (1951).
FR57 U. Fano and G. Racah, *Irreducible Tensors* (Academic Press, Inc., New York).
H54 W. Heitler, *The Quantum Theory of Radiation* (Clarendon Press, Oxford, 1954), third edition.
IZ51 J. Irving and R. Zwanzig, J. Chem. Phys. **19**, 1173 (1951).
K55 T. Kotani, Progr. Theoret. Phys. **14**, 379 (1955).
L50 S. P. Lloyd, Phys. Rev. **80**, 118 (1950).
L51 B. A. Lippmann, Phys. Rev. **81**, 162 (1951).
M48 J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 100 (1948).
N27 J. von Neumann, Göttinger Nachr. 245 and 273 (1927).
N55 J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
S50 J. A. Spiers, Phys. Rev. **80**, 491 (1950).
S52 G. G. Stokes, Proc. Cambridge Phil. Soc. **9**, 399 (1852).
S55 R. L. Stratonovich, J. Exptl. Theoret. Phys. U.S.S.R. **28**, 547 (1955).
T38 R. C. Tolman, *The Principles of Statistical Mechanics* (Clarendon Press, Oxford, 1938).
TG49 H. A. Tolhoek and S. R. de Groot, Physica **15**, 833 (1949).
TG51 H. A. Tolhoek and S. R. de Groot, Phys. Rev. **83**, 189 (1951).
VH54 L. Van Hove, Phys. Rev. **95**, 249 (1954).
VH55 L. Van Hove, Physica **21**, 517 (1955).
W31 H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, New York, 1931).
W32 E. Wigner, Phys. Rev. **40**, 749 (1932).
W53 L. Wolfenstein, Phys. Rev. **92**, 123 (1953).
W54a G. C. Wick, Phys. Rev. **94**, 1228 (1954).
W54b L. Wolfenstein, Phys. Rev. **96**, 164 (1954).
WA52 L. Wolfenstein and J. Ashkin, Phys. Rev. **85**, 947 (1952).
WWW52 Wick, Wightman, and Wigner, Phys. Rev. **88**, 101 (1952).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 1 JANUARY, 1957

Line Shape

R. G. BREENE, JR.*

Geophysics Research Directorate, Air Force Cambridge Research Center, Air Research And Development Command, L. G. Hanscom Field, Bedford, Massachusetts

I. EARLY LINE BROADENING THEORY

IN approaching the subject of line broadening the original work of Michelson provides an entree which utilizes the historical and phenomenological approach simultaneously. The causes of line broadening advanced in this author's early work remain as valid today as they were sixty years ago. So let us therefore utilize this work as our starting point.

A. The Causes of Line Broadening (1895)

Michelson began his consideration of line broadening effects by a summarization of the hypotheses which had been advanced previous to that time to account for these phenomena, and a verbatim restatement may be of interest here.

"1. As a consequence of Kirchhoff's law 'the ratio of brightness of two immediately contiguous portions of a discontinuous bright-line spectrum constantly decreases, if the number of luminous strata is multiplied or if the coefficient of absorption of the single stratus is increased, until the value is reached which, for the same wavelength and the same temperature, corresponds to the ratio in the continuous spectrum of a body completely opaque for a given thickness.[100]†

"2. The direct modification of the period of the vibrating atoms in consequence of presence (sic) of the neighboring molecules.

"3. The exponential diminution in amplitude of the vibrations due to communication of energy to the surrounding medium or to other causes.

"4. The change in wavelength due to the Doppler effect of the component of the velocity of the vibrating atom in the line of sight."

To these causes Michelson added:

"5. The limitation of the number of regular vibrations by more or less abrupt changes of phase amplitude or plane of vibration caused by collisions.

"6. The possible variations in the properties of the atoms within such narrow limits as to escape detection by other than spectroscopic observations."[67]

Having presented the historical list, let us use it in a phenomenological consideration of the general line broadening situation.

* Now at Special Defense Projects Department, General Electric Company, Philadelphia, Pennsylvania.

† References are listed at the end of this article.

We consider an emitting atom which we shall proceed to remove to infinity and reduce the "temperature" to the point where, classically at least, no translational motion exists. Now from the classical picture of a vibrating electron or the simple picture of a pair of energy levels between which our radiating transition takes place, we should expect these conditions to yield a spectral line of a single frequency. We, of course, do not obtain this result, but, rather, we obtain the familiar natural line shape which is attributable to Michelson's Cause 3. We shall not specifically deal with the problem of natural line shape.

Having observed the behavior of the emitted radiation when the atom is removed from all neighbors, its translation motion reduced, let us now give the atom a certain translational velocity. As we are, of course, aware, the resulting atomic "heat" motion leads to a line broadening which Rayleigh first demonstrated as due to the Doppler effect.[73]

B. The Doppler Effect in Line Broadening

First, let us define the word "intensity" of emission (absorption) as the energy emitted (absorbed) per unit area per unit time. Then, if one makes a plot of frequency *vs* intensity for this "almost homogeneous" radiation, one obtains a distribution of intensities over a small frequency range giving rise to a spectral line of definite shape. We are desirous of eventually obtaining the intensity distribution.

We suppose our emitting molecule possessed of a velocity component ξ in the line of sight of the observer. Then if the molecule is emitting radiation of wavelength λ, the Doppler effect decrees that our observer will receive radiation of wavelength λ', such that

$$\lambda' \doteq \lambda(1-\xi/c) \qquad \text{(I.1)}$$

in which expression we assume that $\xi \ll c$.

The distribution of the velocities in the line of sight is given by the Maxwellian distribution as $\exp[-\beta\xi^2]$ where

$$\beta = m/2kT. \qquad \text{(I.2)}$$

In Eq. (I.2) m is the molecular mass, k is the Boltzmann constant, and T is the temperature in degrees Kelvin.

From Eq. (I.1)

$$\xi^2 = \lambda^2(\Delta\nu)^2. \qquad \text{(I.3)}$$

In Eq. (I.3) λ may now be considered as the wavelength corresponding to the maximum intensity of the line (line center), and $\Delta\nu$ as the frequency increment between the line center and the frequency whose displacement corresponds to the line of sight velocity ξ.

Thus, if one arbitrarily equates the line center intensity to one, the Doppler effect alone produces a distribution of intensities over the spectral line which may be represented by the expression

$$I=\exp\left[-\frac{m\lambda^2(\Delta\nu)^2}{kT}\right]. \tag{I.4}$$

Let us define the quantity "half-width" which we shall designate by the symbol δ and which we shall encounter rather frequently throughout our considerations of line shape. Quite simply, this quantity is defined as the width—a frequency measurement—of the spectral line at an intensity equivalent to one-half the maximum intensity in the line. In the present case then

$$\delta=\frac{2}{\lambda}\left(\frac{kT}{m}\ln 2\right)^{\frac{1}{2}}. \tag{I.5}$$

These equations are Lord Rayleigh's solution to Michelson's Point 4 and our explanation of the results of "warming" our atom.

We next bring this emitter in from infinity so that it is surrounded by other atoms and inquire as to the effect of these neighbors on the radiation which is emitted. The answer to this inquiry is furnished by Michelson's Causes 2 and 5. Indeed these line broadening causes are quite as valid from the contemporary viewpoint as from any earlier one.

Cause 2 provides an explanation of the statistical theory which we shall consider in a good bit more detail at a later point. For the statistical considerations we shall find it necessary to proceed from a classical consideration of electron vibrational frequency disturbance to atomic energy level distortion, but the analogy is quite precise.

Cause 5 underlies what has variously been called collision broadening, impact broadening, velocity broadening, and what have you. We shall designate this phenomenon interruption broadening, the mathematics of which were originally contributed by Michelson.

C. The Application of the Fourier Analysis to Line Broadening

In considering the Michelson development, we shall begin by referring to Fig. 1. The circles in the figure in question represent the broadening atoms while the arrow we suppose to be the path of the emitter amongst this assemblage of broadeners. Now Michelson used a billiard ball atomic model so that a collision between the emitter and a broadener will have taken place when the separation of the emitter and broadener centers is equal to or less than their mean atomic diameter. Let us suppose that at point "c" on our emitter path such a collision has just taken place. Then Michelson supposed that this collision has the following effect.

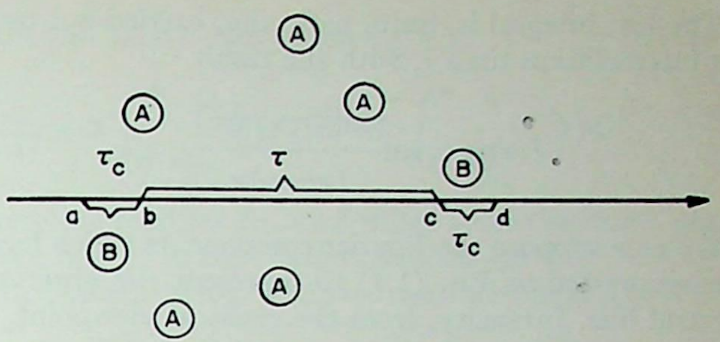


FIG. 1. A physical model for the interruption theory.

For all practical purposes we may suppose that the classical oscillator model of the emitting atom began emitting at time minus infinity and will continue to do so until plus infinlty. The collision which occurred at "c" has the effect of rendering the radiation emitted prior to the collision completely incoherent with the radiation emitted subsequent to this time. This is the crux of the argument for it forms the basis for the Michelson broadening mechanism.

Thus, according to our theory, since a collision has just been undergone at point "b" (on Fig. 1), radiation will effectively begin at this point. If τ is now the time which elapses before the next collision, this is also the time during which radiation continues. Finally the radiation is terminated at point "c," and we are left with a wave train of finite length $c\tau$ where c is, of course, the velocity of light.

Now we may recall that the functions, $e^{i\omega t}$ form a complete orthonormal set, in terms of the members of which any function may be expanded. Further, each of these functions represents a wave train of appropriate frequency and infinite extent. Thus, if we expand our finite wave train in terms of these functions we are essentially expressing the finite wave train as a sum of an infinite number of wave trains of infinite extension. In this expansion the amounts of each train which we take are regulated by the magnitude of the vibrational amplitude. As a result of all this, we obtain

$$M(t)=\int J(\omega)e^{i\omega t}d\omega \tag{I.6a}$$

which is, of course, the familiar Fourier expansion with amplitudes given by

$$J(\omega)=\int M(t)e^{-i\omega t}dt=\text{const}\int e^{i(\omega_0-\omega)t}dt, \tag{I.6b}$$

where ω_0 is the angular frequency of the cut-off wave train.

The last integral is, quite naturally, carried out over the intercollision time τ, with the result

$$J(\nu)=\text{const}\frac{\sin[(\nu-\nu_0)\tau]}{(\nu-\nu_0)\pi}. \tag{I.7}$$

We now suppose the Fourier components which have been expressed as Eq. (1.7) to represent the observed spectral line. Intensity, from the classical viewpoint, is always associated with the square of the amplitude, so that Michelson supposed the distribution of intensities within the line to be given by

$$I=\text{const}\frac{\sin^2[(\nu-\nu_0)\tau]}{(\nu-\nu_0)^2\pi^2}. \tag{I.8‡}$$

If we now suppose the width of the spectral line to be the separation of the minima to either side of the line intensity maximum, the width is

$$\frac{1}{l}\cdot\frac{v}{c}=\frac{\pi v N\sigma^2}{c}. \tag{I.9}$$

Michelson found that Eq. (1.9) gave an order of magnitude agreement with experimental data. Somewhat better agreement was achieved by him when he took the "half-width" as given by Eq. (1.9) and added it to the half-width Eq. (1.5). As Michelson noted the problem is not herewith solved, but it will be of interest to see if ". . . such able contributions . . . recently . . . justify the prediction that a complete and satisfactory theory will be forthcoming in the near future."[67]

Finally, Point 1 is of no particular interest to us, and Point 6, although it shrewdly predicts isotopic spectra, is related to a "pseudobroadening" with which we shall not concern ourselves.

It seems particularly interesting to note the correctness of the factors which Michelson listed as causing line broadening, correctness at least from the contemporary viewpoint.

Certainly of great historical interest is the Lorentz theory of line broadening which appeared some ten years after Michelson's considerations. For our purposes, however, suffice it to say that this theory considered the mechanics of absorption in obtaining an interruption type result.[53]

II. STARK BROADENING

In a study of the line broadening problem that portion of the work done in the field subsequent to the period which we have designated "early" attains such voluminous proportion that some subdivision of the problem must be attempted. We shall divide the field into areas dependent on the phenomenon involved, at least insofar as this is practicable. Our areas of consideration will then be: (1) Stark broadening by the electric fields of the perturbing atoms. (2) Statistical broadening. (3) Interruption broadening. (4) Resonance broadening by like perturbers. (5) Molecular broadening. We shall take up these sometimes overlapping areas of consideration in the order in which we have enumerated them.

Holtsmark's Stark broadening treatment is admittedly a statistical one, but the phenomenon with which his treatment deals is sufficiently unique and, perhaps more important, the subsequent treatments of the phenomenon are sufficiently removed from anything statistical that the assignment of his work to this section seems the proper one.

In the theoretical treatment of the Stark broadening there have essentially been three general treatments each of which remains of importance within its range of applicability. The first of these is the venerable Holtsmark classical treatment of the ionic Stark broadening. The second is Spitzer's quantum treatment of the same problem. The last is the quantum treatment of Stark broadening by electronic fields. We shall consider these three theories in precisely this order.

A. The Simple Holtsmark Theory

Following Lorentz's treatment of broadening by atoms other than the emitter almost fifteen years were to elapse before another detailed study of the problem was to be carried out. This study was, of course, Holtsmark's classical consideration of the broadening by ionic fields.

As we are presently well aware, the Michelson-Lorentz theory of line broadening falls far short of explaining the various shifts which spectral lines may be made to undergo and the shapes which they may be made to assume. Now Stark[101] himself had suggested that some forms of line broadening might be due to fields of the surrounding atoms. Holtsmark[28,29] developed the idea and carried out the analytical attack which demonstrated precisely how these fields might be responsible for this broadening. To this attack we now turn our attention.

We begin by dividing the atoms which surround our emitter and cause broadening into (1) ions, (2) dipoles, or (3) quadrupoles and further specify that there will be only one type of field producer present. A certain spatial configuration of these surrounding, broadening atoms will produce a certain electric field at the emitter. Since all field producers will be of the same type, the only difference between one electric field and another will be a difference in perturber configuration. This is simply because the other factor which enters into the field strength determination, the electrical constants (ionic charge, dipole moment, quadrupole constant) are just that, constants. Now the heat motion of these perturbers will result in various electric fields at the emitter which will occur, of course, with various probabilities. If we

‡ We shall see that the mere application of a random distribution of intercollision times to this intensity distribution yields the more familiar Interruption shape.

suppose a random spatial distribution, then the probability of a particular field strength will be a function of the spatial configuration and hence expressible as a function of the field strength itself. Finally a given field strength will be responsible for a certain Stark splitting of the emitted spectral line. Since the field strength probability is a function of the field strength, the probability of a Stark splitting appropriate to this field strength will also be a function of the field strength. This line of reasonlng and the probability expression obtainable because of it are all very well, but the mechanics of the broadening remains to be particularized.

We now suppose our spectral line to be split into its Stark components by a linear Stark effect due to one of these fields of magnltude, say, F_0. We agree that the distribution of intensities in the split spectral line is given by $I(F_0,\nu)$. If our electric field were F_1, we should obtain a different intensity distribution in the split line, and so on. The spectral line which we actually consider is made up of contributions from all these possible intensity distributions, each distribution weighted by the probability of occurrence of the electric field responsible for the contribution. Then in order to obtain the intensity distribution in the Stark broadened line we must sum over all such distributions weighted by the probability of the distribution occurrence

$$Id\nu=\int_0^\infty I(F,\nu)W(F)dF, \qquad \text{(II.1)}$$

where the $W(F)$ are the field strength probability functions which we shall consider later.

The probability expression of Eq. (II.1) we can calculate but the expression for the intensity in the split line $I(F,\nu)$ is something else again. The actual expression is an extremely complicated one within the limits of the outermost Stark components, but an approximation, which proves quite satisfactory, does exist. Quite simply we suppose $I(F,\nu)$ to be a constant within the limits of the outermost Stark components and zero outside these components. A rectangle of height $I(F,\nu)$ and width $2\nu_m$—wherein we suppose different F values to yield different heights and widths but the same area, $g=2\nu_m\cdot I(F,\nu)$—then results.

This rectangular assumption would appear to require a little change in our integral limits in Eq. (II.1). Let us suppose ourselves interested in a frequency ν on the final spectral line. (This must not be confused with what might be called the initial spectral line, that is, our rectangle.) Then only those rectangles will contribute to the intensity at this frequency for which $\nu_m \geqslant \nu$ while rectangles of lesser ν_m will not contribute. Thus, for a given frequency in the final spectral line, we integrate over all electric field strengths greater than the field strength producing a splitting equal to this frequency. Equation (II.1) then becomes

$$Id\nu=d\nu\int_{F'}^\infty \frac{g}{2\nu_m}W(F)dF=d\nu\int_{F'}^\infty \frac{g}{aF}W(F)dF, \qquad \text{(II.2)}$$

where the relation $2\nu_m=aF$ arises from the linear Stark effect, and F' is that field for which $\nu_m=\nu$.

In evaluating Eq. (II.2) we must first evaluate $W(F)$.

B. Probability of an Electric Field Strength at the Emitter

Holtsmark[29] began with the assumption that the probability of a particular field strength F_0 is a function of F_0, and set out to determine this probability.

Let the components of the field strength at 0 due to the nth perturber be X_n, Y_n, Z_n. The components of F_0 will then be given by

$$X_0=\sum_{n=1}^{N'} X_n; \quad Y_0=\sum_{n=1}^{N'} Y_n; \quad Z_0=\sum_{n=1}^{N'} Z_n. \qquad \text{(II.3)}$$

We desire the probability that X_0 lies between X_0 and X_0+dX_0, etc.

Let the position of the nth broadener be specified by the σ coordinates $x_{1n}, x_{2n}, \cdots, x_{\sigma n}$. Then the X_n, Y_n, Z_n are functions of these σ coordinates. It follows that the X_0, Y_0, Z_0 of Eq. (II.1) are functions of the $N'\sigma$ coordinates thus introduced. An $N'\sigma$ dimensional space is next set up, and to each point in this space there now corresponds a stipulated field strength. These points will be distributed in space according to some probability law, or what amounts to the same thing, some density function.

The probability that all $N'\sigma$ coordinates lie in the range $dx_{11}, dx_{21}, \cdots, dx_{\sigma N'}$ is

$$w_1\cdots w_{N'}dx_{11}dx_{12}\cdots dx_{1\sigma}dx_{21}\cdots dx_{N'\sigma}. \qquad \text{(II.4)}$$

It should be noted that this coordinate distribution establishes our electric field strength within a certain small range.

We integrate Eq. (II.4) over those portions of space satisfying the requirements X_0 between X_0 and X_0+dX_0, etc. Subsequently the proper Dirichlet factors are introduced, and we obtain

$$W(X_0Y_0Z_0)dX_0dY_0dZ_0$$

$$=\frac{1}{8\pi^3}dX_0dY_0dZ_0\iiint_{-\infty}^{+\infty} d\xi d\eta d\zeta\, e^{i(\xi X_0+\eta Y_0+\zeta Z_0)}$$

$$\times\frac{1}{V_{N'}}\int\cdots\int_{\text{all space}} w_1\cdots w_{N'}$$

$$\times e^{i(\xi\Sigma X_n+\eta\Sigma Y_n+\zeta\Sigma Z_n)}dx_{11}\cdots dx_{\sigma N'}. \qquad \text{(II.5)}$$

The X_n, Y_n, and Z_n are functions of the $x_{1n}\cdots x_{\sigma n}$ while the X_0, Y_0, Z_0 are constants in all the integrations.

It should be stipulated now that all the broadening atoms are considered the same, that is, the Stark broadening is not treated for the case where, for atoms, more than one element acts as a broadening agent. In this event, identical electric field strength components X_n, Y_n, Z_n will be the same functions of $x_{1n}, x_{2n}, \cdots, x_{\sigma n}$. It follows from these considerations that for identical X_n, Y_n, Z_n, $w_1, w_2, \cdots, w_{N'}$ will all be equal. As a consequence the inner integral in Eq. (II.5) can be broken up into a product of N' identical integrals, each of which is integrated only over the σ coordinates of a single atom. The result is

$$W(X_0Y_0Z_0)$$

$$=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\cdots\int d\xi d\eta d\zeta e^{-i(\xi X_0+\eta Y_0+\zeta Z_0)}\mathfrak{F}^{N'} \quad \text{(II.6a)}$$

where

$$\mathfrak{F}^{N'}=\frac{1}{V^{N'}}$$

$$\times\left[\int_{\text{all space}}\cdots\int e^{i(\xi X+\eta Y+\zeta Z)}\sigma dx_1dx_2\cdots dx_\sigma\right]^{N'}. \quad \text{(II.6b)}$$

Subsequent to assuming three types of broadeners—that is to say, (1) ions, (2) dipoles, and (3) quadrupoles—Holtsmark carried out a rather lengthy calculation (which we shall not detail) for the field strength probability functions. The result for the dipole is

$$W(X_0Y_0Z_0)=\frac{1}{\pi^3}\frac{c_3}{(c_3^2+F_0^2)^2}. \quad \text{(II.7)}$$

We note that $c_3=4.54\mu N$ where μ is the dipole moment of a broadener.

A "Normalfeldstarke" may be defined as

$$F_n=c_3=4.54\mu N \quad \text{(II.8a)}$$

with β then defined as

$$\beta=F_0/F_n. \quad \text{(II.8b)}$$

Now we replace the volume element $dX_0dY_0dZ_0$ by $4\pi^2F_0^2dF_0$ and Eq. (II.7) results in

$$W(F_0)dF_0=4\pi F_0^2W(X_0Y_0Z_0)dF_0$$

$$=\frac{4}{\pi}\frac{1}{(1+\beta^2)^2}\frac{F_0^2}{c_3^3}dF_0=\frac{4}{\pi}\frac{\beta^2d\beta}{(1+\beta^2)^2}. \quad \text{(II.9a)}$$

Figure 2 furnishes a plot of $W(F_0)$ *vs* β as given by Eq. (II.9a). After we have obtained $W(F_0)$ for the ion and the quadrupole, we can consider the significance of these plots.

In the case of the ion

$$\frac{1}{\beta^{3/2}}=\frac{c_2}{F_0^{3/2}}=\left(\frac{F_n}{F_0}\right)^{3/2}\Longrightarrow F_n=c_2^{2/3}=2.61N^{2/3}\epsilon.$$

If we take as the mean separation of two of the particles $r_0(4/3)\pi r_0^3N=1$ where N is, as usual, the particle density, then this mean particle separation would give about the same result for a normal field strength, namely, for $F'=2.60N^{2/3}\epsilon$.

The ion result is

$$W(F_0)dF_0=W(X_0Y_0Z_0)4\pi F_0^2dF_0$$

$$=\frac{4}{3\pi}\beta^2d\beta[1-0.462\beta^2+0.1227\beta^4$$

$$-0.02325\beta^6+\cdots]. \quad \text{(II.9b)}$$

Equation (II.9b) converges only for small β. For large β one obtains

$$W(F_0)dF_0=\frac{d\beta}{\beta^{5/2}}1.496\left[1+\frac{2.555}{\beta^{3/2}}\right.$$

$$\left.+\frac{14.43}{\beta^3}+\frac{0}{\beta^{9/2}}+\cdots\right]. \quad \text{(II.9c)§}$$

The plot resulting from Holtsmark's versions of Eqs. (II.9b) and (II.9c) is given in Fig. 2.

Finally the probability of a field strength F_0 is required for the quadrupole and has been found as

$$W(F_0)dF_0=\frac{4}{\pi}\frac{d\beta}{\beta^{7/4}}0.805\left[1-\frac{0.730}{\beta^{3/4}}\right.$$

$$\left.-\frac{0.328}{\beta^{3/2}}+\frac{0.621}{\beta^{9/4}}-\frac{0.163}{\beta^3}+\cdots\right] \quad \text{(II.9d)}$$

for small values of β.

Again $\beta=F_0/F_n$ where $F_n=c_4^{3/4}$.

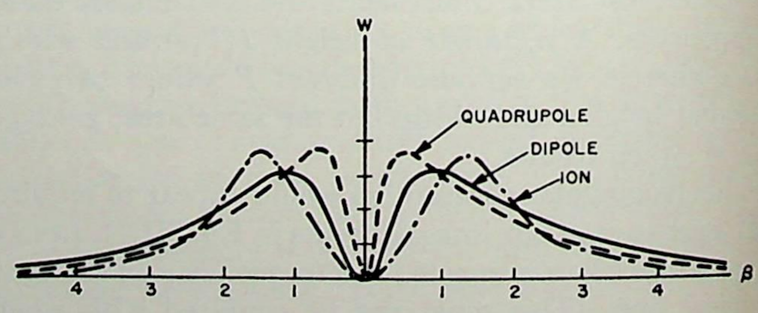


FIG. 2. The field strength probability function for the three cases. (After Holtsmark.[29])

§ Holtsmark obtained

$$W(F_0)dF_0=\frac{d\beta}{\pi\beta^{5/2}}2.350\left[1+\frac{5.106}{\beta^{3/2}}-\frac{7.4375}{\beta^3}+\cdots\right],$$

but a check of the work preceding is sufficient to show that this is not correct. Also see Verweiz[86] as reported by Unsold[97] which is almost correct.

For $0<\beta<1$, Holtsmark obtained

$$W(F_0)dF_0=\frac{4}{\pi}\beta^2 d\beta\frac{4}{3}[1-2.44\beta^2+11.25\beta^4-72\beta^6+\cdots]. \quad \text{(II.9e)}$$

Equations (II.9d) and (II.9e) yield the quadrupole curve in Fig. 2. Figure 2 tells us that, since the probability of zero field strength is zero, we must expect some broadening due to the Stark effect under any conditions. The maxima of the three curves give us the most probable field strengths in the three cases, and it may be noted that the slopes of these probability curves vary. The positions of the maxima indicate that the ion yields the greatest relative broadening, the dipole the intermediate, and the quadrupole the least. It should be recalled, however, that, since β depends on a different "Normalfeldstarke" F_0 in each case, these comparative considerations cannot be directly carried over to comparisons of broadened lines in the three cases.

Equations (II.9) now give the probability $W(F_0)$ for the existence at a certain time of the field F_0 due to one of three causes, which is to be utilized in Eq. (II.2).

C. Line Shape and Half-Widths According to the Early Stark Theory

When Eq. (II.9a) is substituted into Eq. (II.2), the result is

$$Id\nu=d\nu\frac{4}{\pi}\int_{\beta'}^{\infty}\frac{g}{aF_n}\frac{\beta d\beta}{(1+\beta^2)^2}=d\nu\frac{g}{2aF_n}\frac{4}{\pi}\frac{1}{(1+\beta'^2)} \quad \text{(II.10)}$$

in which $\beta'=2(\nu-\nu_0)/aF_n=2\Delta\nu/aF_n$, so that

$$I(\nu)=\frac{2g}{\pi}\frac{aF_n}{a^2F_n{}^2+4(\nu-\nu_0)^2}. \quad \text{(II.11a)}$$

The half-width of the line can be seen to be

$$\delta=aF_n=a4.54N\mu. \quad \text{(II.11b)}$$

The dependence of the line half-width on the electric dipole moment and the gas density is apparent from Eq. (II.11b). With which comment we turn our attention to the ion and the quadrupole.

Equation (II.2) may be rewritten as

$$2\nu_m=aF_n\beta \quad \text{(II.12)}$$

$$Id\nu=d\nu\frac{g}{aF_n}\int_{\beta'}^{\infty}\frac{W(\beta)d\beta}{\beta}. \quad \text{(II.13)}$$

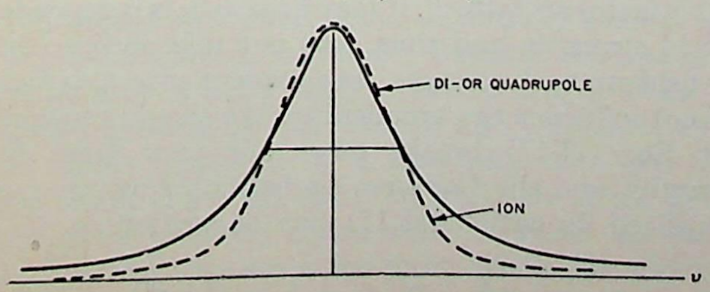


FIG. 3. Spectral line shape as predicted by the Holtsmark Stark theory. (After Holtsmark.[28])

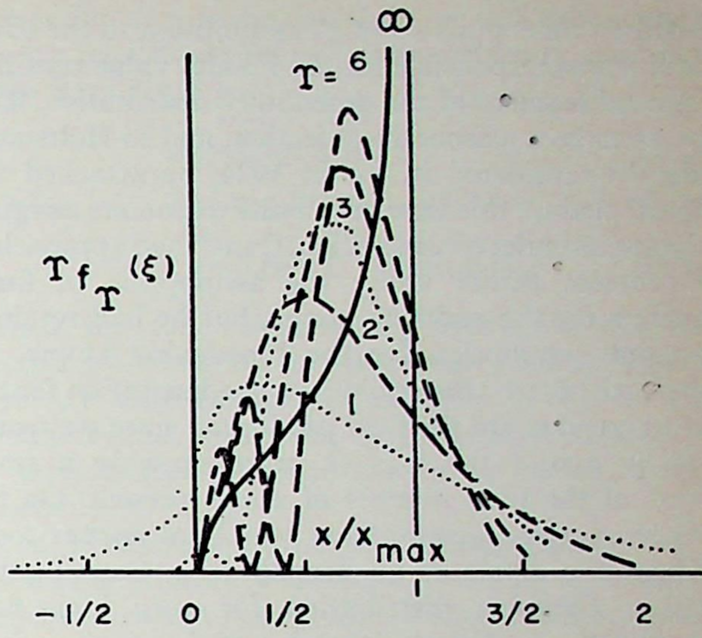


FIG. 4. The line shape according to the Spitzer theory for single encounters as a function of phase shift. (After Spitzer.[79])

Holtsmark evaluated Eq. (II.3) graphically for the cases of the ion and the quadrupole. Figure 3, represents his results for the intensity distribution for all three cases. In Fig. 3 the abscissae have been changed for the three perturbers so that the half-widths coincide, thus giving a curve shape comparison. The half-intensity β's for the ion and the quadrupole may be obtained from the curve as 1.25 and 0.67, respectively. These values of β yield the half-widths:

$$\text{Ion:} \quad \delta=1.25aF_n=3.25aN^{\frac{2}{3}} \quad \text{(II.14a)}$$

$$\text{Quadrupole:} \quad \delta=0.67aF_n=5.53aN^{\frac{4}{3}}A. \quad \text{(II.14b)}\|$$

Equations (II.14a), (II.14b), and (II.11b) show the dependence of the half-width on the electrical properties of the perturbing molecule through ϵ (charge), μ (dipole moment), or A (quadrupole constant) and also on the gas density through N.

D. The Introduction of Finite Molecular Diameters

It may be recalled that we have essentially considered the atoms involved in the theory as points, in that, when the integration was carried out over the atomic positions in space, no portions of space were excluded on the basis of previous occupancy. In 1920 Debye[13] com-

‖ Debye[12] obtained approximate expressions for these field strengths and hence, an idea as to the behavior of the half-widths as follows: Electric charge is taken as 5×10^{-10} $g^{\frac{1}{2}}$-$cm^{\frac{3}{2}}$-sec^{-1} and the radius of the molecule as 10^{-8} cm. We assume that (a) the ion has charge 5×10^{-10} $g^{\frac{1}{2}}$-$cm^{\frac{3}{2}}$-sec^{-1}, (b) the moment of the dipole is 5×10^{-18} $g^{\frac{1}{2}}$-$cm^{5/2}$ sec^{-1} and (c) the quadrupole constant is 5×10^{-26} $g^{\frac{1}{2}}$-$cm^{7/2}$-sec^{-1}. Since the electric field has the units $g^{\frac{1}{2}}$-$cm^{\frac{1}{2}}$-sec^{-1}, and since we may assume the field to be the product of either (a), (b), or (c) and some power of N (molecules-cm^{-3}), thus yielding approximate expressions for the electric field in each of the three cases. Hence, we obtain for (a) $F=\epsilon N^{\frac{2}{3}}$, for (b) $F=\mu N$, and for (c) $F=\theta N^{4/3}$ or 4500 esu, 135 esu, and 4 esu, respectively. Multiplicative constants would yield the values obtained by Holtsmark.

mented on this "punktformig" assumption to the effect that this was responsible for the infinite value resulting for second moment of the probability distribution. This appears to be a reasonable objection, and so Holtsmark evidently considered it, for in 1924, he attacked the problem again,[30] this time with finite diameters assigned to the atoms under consideration. Gans[21] had approached the problem earlier under the assumption of finite diameters for the emitting atoms, but he had retained the point assumption for the broadening atoms. As Holtsmark noted, this would be a good assertion for the case where ions are the field producers, since we would normally expect the ions to make up only a small portion of the total number of atoms present. On the other hand, these assumptions would not appear to be valid for the dipole or quadrupole case. Gans further found a Gaussian distribution for very high field strengths, ". . . wie zu erwarten war."

In this development, Holtsmark retained the simplifying assumption which classifies the broadening atoms as ions, dipoles, or quadrupoles, that is, he again took only the first term in the series for the potential. Since these assumptions are predicated on large R (radius), it is apparent that when the gas density or pressure is high, R no longer remains large enough to justify them. How high "high" is, is, of course, a matter for discussion.¶ For these "high" densities the calculations cannot be carried out, but the field strength distribution in these cases is assumed Gaussian. Holtsmark thus limits himself to those gas densities where the first term in the potential series does give a good approximation of the electric field.

We begin with Eq. (II.5) rewritten in a slightly different form.

$$W(F_0)=\frac{2F_0}{\pi}\int_0^\infty dss\sin(F_0s)\frac{G(s)}{Q(s)}, \quad \text{(II.15a)}$$

where

$$G(s)=\int d\sigma we^{i\Sigma(\xi X_n+\eta Y_n+\zeta Z_n)} \quad \text{(II.15b)}$$

$$Q(s)=\int sd\sigma \quad \text{(II.15c)}$$

and where the integration over the angular coordinates has been carried out.

Now in the computation of $L(s)$ and $M(s)$, the finiteness of the atomic diameters is to be taken into account. In the earlier computation, it was possible to transform Eq. (II.5) into Eqs. (II.6), a product of identical integrals. This was legitimate, due to the independency of the atoms in that the motion of one atomic point is not interfered with by the other atomic points. This integral product is no longer admissible after finite diameters have been assigned the atoms, for we may not now allow the center of an atom to be separated from the center of another atom by less than this atomic diameter. We let

$$\Xi_n=e^{isF_n\cos\chi_n} \quad \text{(II.16a)}$$

so that

$$G(s)=\int_k d\sigma w\prod_{k-1}^{N'}\Xi_n$$

$$=\int_{k_1}\Xi_1w_1d\sigma_1\int_{k_2}\Xi_2w_2d\sigma_2$$

$$\times\int\cdots\int_{k_{N'}}\Xi_{N'}w_{N'}d\sigma_{N'} \quad \text{(II.16b)}$$

where the first integral is carried over the space not occupied by the other $(N'-1)$ atoms and the 0th atom (emitter). Thus, the first integral is dependent on the coordinates of the other atoms and must be included under the integral sign of the second integral, and so on. This fact obviously does not simplify matters. The difficulty may be eliminated, however, by selecting a suitable initial distribution for atoms. This selection would appear to be justified, since our results should not depend on the arbitrary initial distribution of the atoms. As an initial distribution we group all the atoms together at a point, of arbitrary initial selection but remaining fixed during the first $N'-1$ integrations. The position of the first atom is chosen for this union. After locating the atoms in this manner, we may move atom N' about space in the process of the integration, while keeping the remaining atoms at position one. This process is carried out for $N'-1$ of the atoms. Without belaboring the details of the remainder of the calculation, let us remark that Holtsmark obtained

$$W(F_0)=\left[\frac{4}{\pi}\frac{F_0^2F'}{(F'^2+F_0^2)}\right]e^{a'/3}\left\{1+b\frac{3F'^2-F_0^2}{F'^2+F_0^2}\right\} \quad \text{(II.17)}$$

where $F'=4.54N\mu$.

The factor in brackets in Eq. (II.17) is identical to Eq. (II.9a) the probability for the case of point atoms, while the remaining factors essentially correct it for finite diameters. Gans had already obtained the corrective factor $e^{a'/3}$ which, it may be noted, is independent of field strength, and thus does not take into account any field strength changes. As a consequence, this factor will not influence the broadening. We should remember that Eq. (II.17) breaks down for very large field strengths, and the Gaussian probability function must be utilized. Equation (II.17) may be written as

$$W(\eta)=\frac{4}{\pi}\frac{\eta^2}{(\eta^2+1)^2}e^{a'/3}\left\{1+b\frac{3-\eta^2}{1+\eta^2}\right\}, \quad \text{(II.18)}$$

¶ In a somewhat similar consideration, Spitzer (see infra, this Chapter) takes as a limit a pressure such that $R=10r_0$ where r_0 is the radius of the Bohr orbit.

where

$$\eta = F_0/F'.$$

A simple calculation shows the maximum of the probability curve to fall at $\eta = 1-b$, while this maximum is at $\eta = 1$ for the dipole on the uncorrected curve. The resulting line width is changed by the same percentile amount.

E. Dipole Line Shape from Refined Stark Broadening Theory

Although Holtsmark did not work out the line shape for this corrected case, let us make the short calculation necessary to obtain an idea of this shape. We shall utilize Eq. (II.13) and assume that for $\eta' \leqslant \eta \leqslant \eta''$. Equation (II.18) yields the correct form of $W(\eta)$ while for $\eta'' \leqslant \eta \leqslant \infty$ $\mathrm{Erf}(\eta)$ yields the correct $W(\eta)$. We are assuming η'' very large. Equation (II.13) becomes

$$I(\nu)d\nu = d\nu e^{a'/3}\frac{4}{\pi}\frac{g}{aF_n}\int_{\eta'}^{\eta''}\frac{\eta}{(1+\eta^2)^2}\left\{1+b\frac{3-\eta^2}{(1+\eta^2)}\right\}d\eta$$

$$+d\nu\frac{4}{\pi}\frac{g}{aF_n}\int_{\eta''}^{\infty}\mathrm{Erf}(\eta)d\eta$$

$$=\frac{4}{\pi}\frac{g}{aF_n}\frac{e^{a'/3}}{4}\frac{(2+5b)+(2+4b)\eta'^2}{(1+\eta'^2)^2}+O(\eta''). \quad \text{(II.19)}$$

The spectral line given by Eq. (II.19), if we neglect the $O(\eta'')$ term, will still be symmetrical about the line center. For small b, and b has been assumed small, the curve shape as given by Eq. (II.19) is very nearly identical with the shape given by Eq. (II.10). A good approximation for values of b up to 0.0020 is

$$I(\nu)=\frac{4}{\pi}\frac{g}{2aF_n}e^{a'/3}\frac{1+2b}{(1+\eta'^2)}=\frac{2g}{\pi}e^{a'/3}\frac{(1+2b)aF_n}{a^2F_n{}^2+4(\nu-\nu_0)^2}$$

$$=\frac{g}{\pi}e^{a'/3}\frac{(1+2b)(\delta/2)}{(\nu-\nu_0)^2+(\delta/2)^2}. \quad \text{(II.20)}$$

Holtsmark calculated b for four monatomic and four diatomic gases on the basis of Debye's work[13] and obtained values of $b<0.0020$ in all but one instance.

F. The Line Shift

We might now, with Holtsmark, consider another form of broadening which may arise. Suppose the line is not split to give a band for the field strength F as was assumed in deriving Eqs. (II.10) and (II.20), but that it is simply displaced by a frequency $\Delta\nu$. Now for a linear Stark effect the right (or left) side of the $W(F_0)$ curve is the intensity distribution curve for the broadened spectral line, since the probabilities for a shift of the line by an amount corresponding to this field strength, and, hence, the relative intensities at those frequencies. If γ is the line shift per unit field strength, we may read the shifts of the line intensity maxima directly from Fig. 4 as

$$\Delta\nu = 1.5\gamma F_n \text{ Ion} \quad \text{(II.21a)}$$

$$\Delta\nu = 1.0\gamma F_n \text{ Dipole} \quad \text{(II.21b)}$$

$$\Delta\nu = 0.65\gamma F_n \text{ Quadrupole.} \quad \text{(II.21c)}$$

Thus, this type consideration gives an asymmetrical spectral line whose maximum intensity has undergone a frequency shift. Qualitative examples of this shift and asymmetry had been given by Takamine and Kokubu[82] for certain He lines and for many metal lines by Takamine.[81]

The Holtsmark Stark broadening theory has its region of applicability although, it is not, of course, the complete answer to the problem. Some conclusions as to applicability were drawn by Holtsmark as a result of his comparison of theory with experiment.

Two possibilities arose from these comparisons. (1) The Stark effect, if it is a factor, is not the only factor which contributes to the broadening of a spectral line. With this statement there can be little disagreement. (2) Simultaneous broadening by both ions and dipoles or quadrupoles occurs and the fields produced may be inhomogeneous ones.

A further comment in definition of this inhomogeneity may be in order. Let us consider, say, an alkali atom in which we may only be concerned with the spectrum-producing valence electron. For our purposes here we can deal either with a Bohr orbit for this electron or those regions of space where $\bar{\psi}\psi d\sigma$ is relatively high for the electron. Then in order that the fields producing Stark broadening be homogeneous for this case, these fields must be essentially constant over the region of the orbit or of high $\bar{\psi}\psi d\sigma$.

We have restricted our gas to low pressure in order that the results may be valid, and another restriction has been inferred but not actually stipulated. If a given field F at time t is to broaden a line into one of the rectangles which have been assumed for the various values of F, the F must be a constant for a time interval sufficient to allow the atom to emit under conditions which will produce this rectangle. The field at the emitter varies due to the motion of the atomic broadeners, and this means that an upper limit, which depends on the lifetime of the upper of the two quantum states involved in the emission, must be placed on these atomic velocities, and hence, the gas temperature.

These considerations also demonstrate the necessity for forbidding the application of this theory to the fields of neighboring electrons.

G. The Adiabatic Quantum Broadening Problem

We first remark that Epstein[16] has shown that, when the problem of the Stark effect in hydrogen is considered

in parabolic coordinates, one obtains

$$E^{(1)}=\frac{3\hbar^2F}{2\mu\epsilon}n(k_2-k_1)=\frac{q}{r^2} \qquad \text{(II.22a)}$$

for the linear Stark effect, and

$$E^{(2)}=-\frac{F^2}{16\mu^2}\left(\frac{\hbar}{\epsilon}\right)^6 \times n^4[17n^2-3(k_2-k_1)-9m^2+19] \qquad \text{(II.22b)}$$

for the quadratic Stark effect.

Spitzer began with three simplifying assumptions: "(1) the matrix elements of the interaction potential between states of different total quantum number may be neglected, and . . . the other matrix elements may be computed on the assumption that the atom is in a homogeneous field of strength $Ze^2/r\cdots$. (2) Each collision . . . may be assumed isolated from all others. (3) The mass of the colliding particles may be taken infinitely large."[79]

Assumption (1) supposes a homogeneous field F over the "boundaries" of the atom, or, say, over the electron orbit. If a field is imposed from "outside" there can be little objection to this. Now, however, we are supposing our field to be produced by an ion, and, if the separation of the hydrogen and the ion r is not large compared to the average distance of the electron from the hydrogen nucleus $\langle r_e\rangle$, this approximation surely breaks down. Thus, some lower limit R must be set on r, and Spitzer chooses this limit as $R>10\langle r_e\rangle$. Now $\langle r_e\rangle$ may have as its maximum value $3na_0^2/2\cdots a_0$ is the radius of the first Bohr orbit—so that $R>8.0\times10^{-8}\ n^2$ where n is the total quantum number. The exchange integrals as well as quadratic Stark effects will be small for these distances as we may show by Eqs. (II.22).

$$\frac{\Delta\nu_1}{\Delta\nu_2}=\frac{E^{(1)}}{E^{(2)}}=\frac{24n(k_1-k_2)R^2}{n^4[17n^2-3(k_2-k_1)^2+9m^2+19]a^2}.$$

When k_2-k_1 is set equal to $n/2$ and m and 19 are neglected, the result is

$$\frac{\Delta\nu_1}{\Delta\nu_2}\doteq\frac{3R^2}{4a^2n^4}=\frac{27}{16}\left(\frac{R}{\langle r_e\rangle}\right)^2>150.$$

This is the binary assumption which, together with the third assumption, will become familiar ones. It is certainly obvious that the treatment of a two particle interaction is far simpler than the treatment of a three or more particle interaction. In addition, it is rather apparent that at the lower pressures the approximation should be a reasonable one by virtue of the following. The highest probability is that the separation of a single one of the surrounding molecules from the emitter is sufficiently small to insure that its interaction overshadows that of the remaining broadeners. It might be said that, although Assumption (3) is reasonable for heavy particles, this will certainly not be true where the ions are electrons.

In considering the broadening problem we suppose the change in the potential as adiabatic so that the quantum state of the system will remain unchanged during a collision. Let us digress for just a moment on the subject of adiabatic processes, since we shall encounter them continuously in our further consideration of line broadening theories.

An adiabatic process may be considered as one in which the entropy of the system—or the degree of disorder thereof—remains unchanged. Now if the probability that an atom is in state X has the value unity before and after the occurrence of some phenomenon which affects the atom "system," the degree of disorder of this system remains unchanged and hence, the occurrence of the phenomenon constitutes an adiabatic process. On the other hand, if the probability for state X is unity at the initiation of the process and changes by virtue of the occurrence of this process to 0.6 for state X, 0.3 for state Y, and 0.1 for state Z, the degree of system disorder has changed, and in consequence, a non-adiabatic process has occurred.

In the present adiabatic case, a two state atom with an upper state A and a lower state B is considered where E_A and E_B, respectively, are the energy of the two states.

Since the electric field is assumed directed along the z axis of the emitting hydrogen atom, and since this electric field is radial from the perturber to the emitter—this, of course, would only be strictly correct for the one perturber assumed—the coordinate system will rotate with the passage of the ion so that the z axis is always directed toward the ion. Adiabaticity requires that this rotation occur for the following reason. The quantum number m specifies the projection of the angular momentum on the electric field of the ion. In order that m remain constant as required by the adiabatic hypothesis the atom must rotate with the changing direction of the ionic field.

We now define $\mathcal{H}$ as the atom-field Hamiltonian, ψ_A as the eigenfunction for the upper state (no photons in the field), and ψ_{b_i} the function for the lower state (a photon of frequency ν_i in the field).

The Schroedinger equations,

$$\mathcal{H}\psi_a^{(0)}=E_A(t)\psi_a(t) \qquad \text{(II.23a)}$$

$$\mathcal{H}\psi_{b_i}^{(0)}=[E_B(t)+h\nu_i]\psi_{b_i}(t) \qquad \text{(II.23b)}$$

result.

In Eqs. (II.23) t is a parameter, and $E_A(t)$ and $E_B(t)$ are the atomic level energies *as perturbed by the linear Stark effect* of the ionic field.

The state growth equations are

$$i\hbar\dot{a}=E_A(t)+a(t)+\sum_i\mathcal{H}_{ab_i}b_i(t) \qquad \text{(II.24a)}$$

$$i\hbar\dot{b}_i=\{E_B(t)+h\nu_i\}b_i(t)+\mathcal{H}_{b_ia}a(t). \qquad \text{(II.24b)}$$

As usual we suppose $a(0)=1$ and $b_i(0)=0$ to obtain the solutions

$$a(t)=e^{-\frac{1}{2}\gamma t}\exp\left\{-\frac{i}{\hbar}\int_0^t E_A(t')dt'\right\} \qquad \text{(II.25a)}$$

$$b_i(t)=-\frac{i\mathcal{H}_{b_i a}}{\hbar}\exp\left\{-2\pi i\nu_i t-\frac{i}{\hbar}\int_0^t E_B(t')dt'\right\}$$

$$\cdot\int_0^t e^{-\frac{1}{2}\gamma t'}dt'\exp\left\{2\pi i\nu_i t'+\frac{i}{\hbar}\right.$$

$$\left.\times\int_0^{t'}[E_B(t'')-E_A(t'')]dt''\right\}. \qquad \text{(II.26b)}$$

By time $t=\infty$ the emitter will certainly be in the ground state, and a photon of frequency ν will be in the field. The intensity distribution in the spectral line is surely given by $|b_{i(\nu)}(\infty)|^2$, since this will represent the probabilities for the appearance of these various frequencies. If we call $I'(\nu)$ the intensity of the frequency ν there results

$$I'(\nu)=\frac{(\mathcal{H}_{b_i a})^2}{\hbar^2}\left|\int_0^\infty e^{-\frac{1}{2}\gamma t'}dt'\exp\left\{2\pi i\nu_i t'+\frac{i}{\hbar}\right.\right.$$

$$\left.\left.\times\int_0^t[E_B(t'')-E_A(t'')]dt''\right\}\right|^2. \qquad \text{(II.26)}$$

A few algebraic manipulations lead to

$$I(x)=\frac{\gamma}{2\pi x^2}\left\{1+\frac{W(\Upsilon,\xi)}{\gamma}\Upsilon^2 f_\Upsilon^2(\xi)\right\}, \qquad \text{(II.27a)}$$

where

$$f_\Upsilon(\xi)=\int_{-\infty}^{+\infty}\frac{\exp[i(\xi u-\Upsilon\tan^{-1}u)]}{1+u^2}du \qquad \text{(II.27b)}$$

and

$$\xi=\frac{xR}{v};\quad u=\frac{vt''}{R};\quad t'=t''+t_0;\quad \Upsilon=\frac{q_A-q_B}{\hbar Rv} \qquad \text{(II.27c)}$$

where q has been defined by Eq. (II.22a). $W(\Upsilon,\xi)=v/R$ is the number of collisions per second.

From Eq. (II.27a) it is possible to show that for ξ small (v large or the density small and hence ρ large) the interruption distribution (which we consider in Chapter IV) holds. It is also possible to demonstrate the applicability of the statistical distribution (which we consider in Chapter III) for ξ large (v small or the density large). Finally, Eq. (II.27a) may be integrated, if rather laboriously—an operation which we shall not detail.

Figure 4 gives the results of Spitzer's calculations for several values of Υ. $x_{\max}$ is nothing more than $(q_A-q_B)/\hbar R^2$. The limiting value for $\Upsilon=\infty$ corresponds, of course, to zero velocity according to Eq. (II.27c). Thus, this value of Υ should yield the line shape as given by the statistical theory.

These then are the Spitzer results for the Stark broadening of a line due to the adiabatic collision of the emitter with a broadener of a particular velocity and distance of closest approach. A detailed consideration of the nonadiabatic case for the first Lyman lines of hydrogen showed that the breakdown of the adiabatic hypothesis (for this particular case) necessitated the replacement of Υ by σ where now $\sigma^2=1+\Upsilon^2$. We might remark that there is no particular reason to expect the same replacement for any other situation since the problem applies solely to these first Lyman lines.

The problem which remains consists of the averaging of Eq. (II.27a) over the requisite distributions of the optical collision diameter (distance of closest approach) and the velocity.[80] Together with the sum over the various type broadeners these operations result in

$$I(x)=\frac{\gamma}{2\pi x^2}\left\{1+\frac{1}{\gamma}\sum_m\int_0^\infty dv\right.$$

$$\left.\times\int_0^s W(\rho,v)\Upsilon^2 f_\sigma^2(\xi)d\rho\right\}. \qquad \text{(II.28)}$$

$W(\rho,v)$ is the collision frequency for collisions of diameter ρ and velocity v. The introduction of a Maxwellian velocity distribution and a random collision diameter distribution leads to an integral whose evaluation requires approximations and graphical evaluations which we shall not detail. Equation (II.28) may be rewritten without much difficulty, however, as

$$I(x)=\frac{\gamma}{2\pi x^2}\left\{1+\frac{4}{\gamma}\left(\frac{\pi}{x}\right)^{\frac{1}{2}}\right.$$

$$\left.\times\sum_m\frac{N_m(q_{A_m}-q_{B_m})^{\frac{1}{2}}}{\hbar^{\frac{1}{2}}}(\gamma_1)^{\frac{1}{2}}H_m(x)\right\} \qquad \text{(II.29a)}$$

where

$$H_m(x)=\int_0^\infty e^{-u}du\int_0^{\gamma_1/\gamma_2\sqrt{u}} f_\sigma^2(\xi)\frac{d\xi}{\xi} \qquad \text{(II.29b)}$$

$$\gamma_1=\frac{(q_{A_m}-q_{B_m})l_m x}{\hbar};\quad \gamma_2=\frac{(q_{A_m}-q_{B_m})(l_m)^{\frac{1}{2}}}{\hbar s};$$

$$l_m=\frac{1}{2kT}\frac{M_B m_H}{M_B+m_H} \qquad \text{(II.29c)}$$

and N_m is the broadener density in the collision class m.

The Holtsmark β of our classical Stark considerations may be obtained by the transformation

$$W(\beta)d\beta=I(x)dx \Longleftrightarrow \beta^2 W(\beta)$$

$$=\frac{3}{\pi^{\frac{1}{2}}}\sum_m\frac{\gamma_2 H_m(x)N_m}{N} \qquad \text{(II.30a)}$$

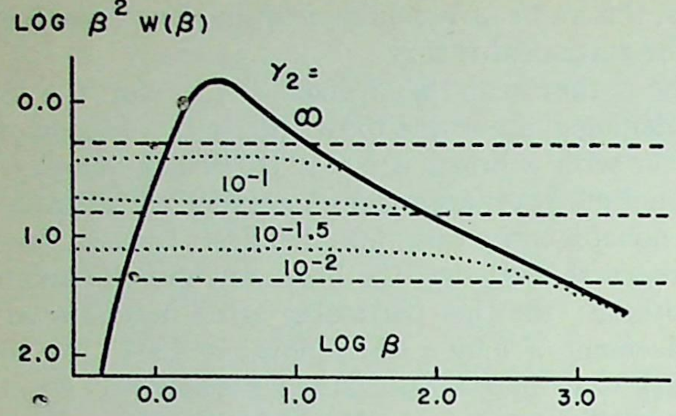


FIG. 5. The line shape for a single H line component. The solid line is the Holtsmark result. The dashed lines arise from the simple interruption theory. (After Spitzer.[80])

where

$$\beta=\frac{xs^2\hbar}{q_A-q_B}. \tag{II.30b}$$

These results through Fig. 5 afford a convenient method of comparing this theory with the interruption and statistical theories as well as demonstrating the predictions of the theory itself.

The horizontal lines in the figure represent the interruption result—the oft-called Weisskopf result—for several values of γ_2. Inner agreement is actually obtained only for $\gamma_2=0.047$. This is a demonstration of the error involved in the failure of the simple interruption theory to include the effect of distant collisions and nonadiabaticity. We mention this fact with no special emphasis, since we shall go into a much more detailed consideration of these facets of the interruption theory in Chapter 4.

One other rather obvious point seems worthy of mention in connection with Fig. 5. Equation (II.29c) tells us that γ_2 is directly proportional to the density, since it is inversely proportional to the mean molecular separation. The result is that as the density increases the theory tends toward the statistical theory, and conversely, as the pressure decreases, the theory tends toward the interruption theory. This may be seen directly from Fig. 5. These are facts which we shall encounter on several subsequent occasions, but this seems a particularly simple way to demonstrate these facts about the two types of theories. One more item of interest may be garnered from a perusal of the figure in question. The parameter γ_2 in addition to being density dependent, is inversely velocity dependent through l_m [see Eq. (II.29c)]. Firstly then, it is rather obvious that we tend toward the statistical theory with decreasing velocity—this too we shall later divine from different considerations.

In toto the quantum Stark theory is restricted by the approximations: (1) linear Stark effect; (2) binary collisions; (3) broadeners of infinite mass. These restrictions tend in general to restrict the applicability to low pressures such that $R>r_e$ and ionic (as opposed to electronic) broadeners.

Spitzer's and Holtsmark's work on ionic broadening still appears to leave the field of electronic Stark broadening rather neglected so that for this area of consideration we must turn our attention to recent work of Kivel, Bloom, and Margenau,[40a] and Meyerott and Margenau.[66a]

H. The Broadening Effect of Electrons

Perhaps the prime reason for the inability of the theories which we have thus far considered to deal with the broadening of spectral lines by electrons is their implicit assumption of a classical path. When an expression for the perturbation due to the broadening particle depends primarily on an emitter-broadener separation, then a classical path assumption has surely been made. Now the Heisenberg uncertainty principle tells us that such an assumption constitutes an extremely poor approximation when the particles involved are electrons, so we must look to a completely quantum treatment of this type broadening.

The familiar time dependent form of the Schroedinger equation

$$\left(\mathcal{H}-i\hbar\frac{\partial}{\partial t}\right)\psi=0 \tag{II.31}$$

we suppose satisfied by the expansion

$$\Psi=\sum_n\sum_\Lambda{}^n a_{n\Lambda}(t)\psi_n(\mathbf{r})\varphi_\Lambda{}^n(\mathbf{R})\times\exp[-i(\mathcal{E}_\Lambda{}^n+E_n)t/\hbar]. \tag{II.32}$$

The Hamiltonian appearing in Eq. (II.31) may be written out as

$$\mathcal{H}=\mathcal{H}_a(\mathbf{r})+\mathcal{H}_e(\mathbf{R})+\mathcal{H}'(\mathbf{r},\mathbf{R})+\mathcal{H}''(\mathbf{r}) \tag{II.33}$$

where $\mathcal{H}_a(\mathbf{r})$ is the Hamiltonian of the isolated atom; $\mathcal{H}_e(\mathbf{R})$ is the Hamiltonian of the electron in the field of this atom; $\mathcal{H}'(\mathbf{r},\mathbf{R})$ the Hamiltonian of the atom-electron interaction; and $\mathcal{H}''(\mathbf{r})$ the Hamiltonian of the atom-field interaction. The eigenfunctions of the expansion Eq. (II.32) are given by

$$\left(\mathcal{H}_a+\int\bar{\varphi}_\Lambda{}''\mathcal{H}'\varphi_\Lambda{}''d\mathbf{R}\right)\psi_n{}^\Lambda=E_n{}^\Lambda\psi_n{}^\Lambda \tag{II.34a}$$

and

$$\left(\mathcal{H}_e+\int\bar{\psi}_n{}^\Lambda\mathcal{H}'\psi_n{}^\Lambda d\mathbf{r}\right)\varphi_\Lambda{}^n=\mathcal{E}_\Lambda{}^n\varphi_n{}^\Lambda. \tag{II.34b}$$

The substitution of Eq. (II.31) into Eq. (II.32) results in the state growth equation

$$i\hbar a_{mM}\exp[-i(\mathcal{E}_M{}^m+E_m)t/\hbar]$$
$$=\sum_n\sum_\Lambda{}^n a_{n\Lambda}\mathcal{H}_{mn}''\int\bar{\varphi}_M{}^m\varphi_\Lambda{}^n d\mathbf{R}\times\exp[-i(\mathcal{E}_\Lambda{}^n+E_n)t/\hbar]$$
$$+\sum_{n\neq m}\sum_\Lambda{}^n a_{n\Lambda}\int\bar{\varphi}_M{}^m\mathcal{H}_{mn}'\varphi_\Lambda{}^n d\mathbf{R}\times\exp[-i(\mathcal{E}_\Lambda{}^n+E_n)t/\hbar] \tag{II.35}$$

wherein we must recall that the electron eigenfunctions are not orthonormal.

The second term on the right of Eq. (II.35) represents what Kivel, Bloom, and Margenau (KBM hereafter) have called the "polarization effect." For the two states of our atom degenerate or nearly so this term is actually a polarization term. For such not the case the combination of these states through this term represents a quenching effect. We shall consider this in a little more detail at a later point, but let us now turn our attention to the subject of this section which is representable by the first term on the right side of Eq. (II.35).

The first term represents what these authors have dubbed "universal broadening," or, perhaps more concisely, it represents the interaction effect responsible for this phenomenon. The effect was termed universal since it will always be present by virtue of requiring no change in the atomic state. It represents a radiating atomic transition during which a change in the interacting electron energy takes place.

In our sketchy treatment of the mathematics of the situation we shall begin by neglecting the $\mathcal{H}_{nm}'$ in Eq. (II.35). We thus consider a two state atom whose transition simply takes place in the field of a broadening electron. The coefficient a_Λ we take to be associated with the upper of the two states wherein no photon is in the radiation field. The coefficient b_{rM} refers to the lower of the two atomic states in conjunction with which a photon of frequency ω_r is in the field. Further $\hbar\Omega_{\Lambda' M}=\mathcal{E}_{\Lambda'}{}^1-\mathcal{E}_M{}^0$ and $\omega_r=\omega_{\text{atom}}-\omega$. Next we shall accept the exponential decay quality of a_Λ, the proof of which may be found in Kivel, Bloom, and Margenau. At any rate, one may then obtain for the line intensity distribution

$$I_U(\omega)=\Upsilon_r\sum_M|b_{rM}(\infty)|^2$$

$$=\Upsilon_r|\mathcal{H}'/\hbar|^2\frac{\sum_M\left|\int\bar{\varphi}_M{}^0\varphi_n{}^1d\mathbf{R}\right|^2}{\gamma^2+(\omega+\Omega_{\Lambda M})^2}. \quad \text{(II.36)}$$

The introduction of the time development matrix representation for $a(t)$ and $b(t)$ and a goodly bit of additional manipulation of an algebraic nature leads to

$$I_U=\frac{\Upsilon|J/\hbar|^2}{\gamma^2+\omega^2}\left\{1+\frac{R}{\gamma^2+\omega^2}\right\} \quad \text{(II.37a)}$$

where

$$R=\sum_\Lambda\Upsilon_{\Lambda\Lambda}{}^0\sum_M R_{\Lambda M};$$

$$\text{(II.37b)}$$

$$R_{\Lambda M}=\frac{A_{\Lambda M}}{\gamma^2+(\omega-\Omega_{\Lambda M})^2}+\frac{B_{\Lambda M}}{\gamma^2+(\omega+\Omega_{\Lambda M})^2}$$

and

$$\gamma\equiv\pi\Upsilon|J/\hbar|^2. \quad \text{(II.37c)}$$

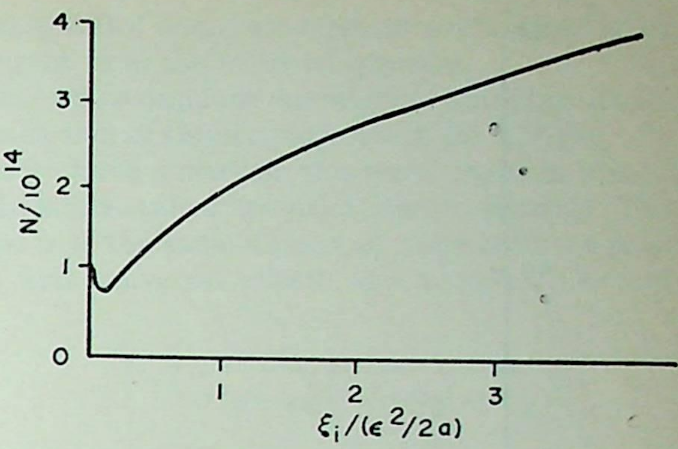


FIG. 6. Electron density *versus* energy. In the region below the curve the universal broadening effect is negligible compared to the natural line width. (Kivel, Bloom, and Margenau.[40a])

In Eq. (II.37b) the first term in $R_{\Lambda M}$ arises for the situation in which energy is not conserved during the radiation process. The resonance condition,

$$\Omega_{\Lambda M}=\omega$$

we suppose to exist. Then, when $\omega_r>\omega_a$ the radiation has received more energy than the atom gave up. It hasn't received this additional energy from the electron, however, since the electron too has gained energy. The effect, resulting from this nonconservative phenomenon, these authors point out, will be small. The second term in the expression for $R_{\Lambda M}$ represents energy transfer which adds up in such a way that conservation can be claimed. This may be demonstrated as was the first case.

The remainder of our consideration of this treatment will be just sufficient so that the terms with which we shall be presented will be defined with a degree of clarity. To begin with the electron-field interaction may be written as

$$\mathcal{H}_{nn}'=-\frac{\epsilon^2}{|\mathbf{R}|}+\epsilon^2\int\frac{|\psi_n(\mathbf{r})|^2}{|\mathbf{R}-\mathbf{r}|}dr.$$

We may surely write the unperturbed electron functions as plane waves

$$u_\Lambda=\frac{1}{\sqrt{V}}\exp(i\mathbf{k}_\Lambda\cdot\mathbf{R})$$

where now $\mathbf{k}_\Lambda$ is the electron momentum vector. At the same time we define a momentum transfer vector $\mathbf{K}=\mathbf{k}_M-\mathbf{k}_\Lambda$ which enters our considerations through the electron eigenfunction products arising in the matrix elements. Also

$$\frac{1}{V}\int e^{i\mathbf{K}\cdot\mathbf{R}}\mathcal{H}_{nn}'dR=(\mathcal{H}_{mn}')_{\Lambda M}=\frac{1}{V}\frac{4\pi\epsilon^2}{K^2}[G_{nn}-1].$$

One finally obtains for the universal broadening line shape

$$I_U=\frac{\Upsilon_r|\mathcal{H}'/\hbar|^2(\gamma+\gamma_U)/\gamma}{\omega^2+(\gamma+\gamma_U)^2}. \quad \text{(II.38)}$$

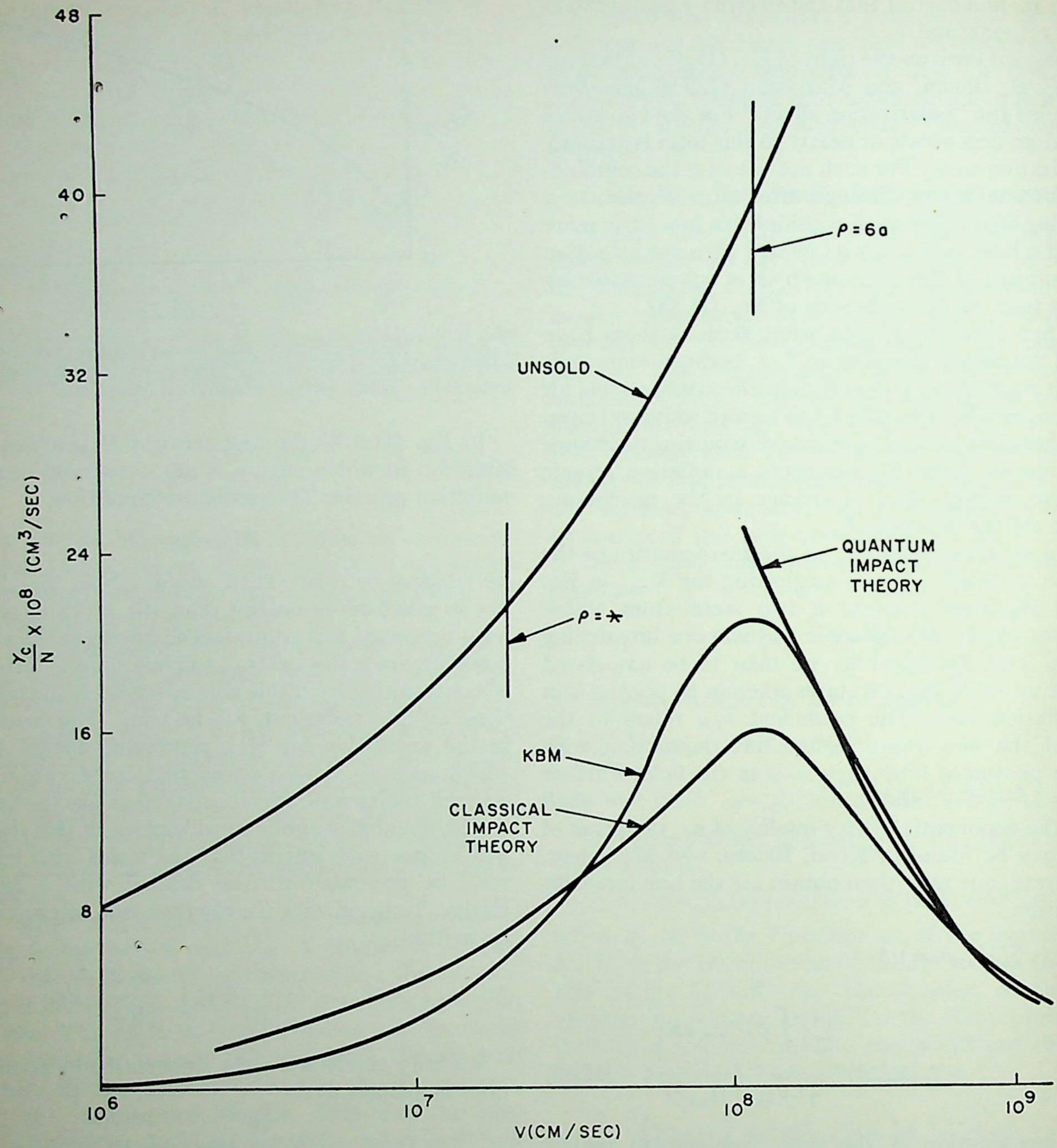


FIG. 7. The collision frequency as a function of velocity according to various of the electron broadening theories. (After Meyerott and Margenau.[66a])

The line width of this distribution is given by

$$\delta = 2\gamma_U = 2fS_U\gamma, \qquad \text{(II.39)}$$

where we now must needs define the additional terms:

$$f = 2nv_\Lambda\sigma_\Lambda/\gamma \Longrightarrow \sigma_\Lambda = \pi(\hbar/mv_\Lambda)^2.$$

In Fig. 6 is given a representation of the region of applicability of this theory. The curve represents a value of one tenth for the ratio γ_U/γ, that is, the ratio between the universal and the natural widths. Below the curve it does not hold due to certain approximations which were made in the expansions for the time development matrices for $a(t)$ and $b(t)$. As we get farther above the curve (very high densities) a statistical type (static) theory may be used.

As Meyerott and Margenau have pointed out, S_U may be approximately written as

$$S_U = 4.3(1 - e^{-4.1x}). \qquad \text{(II.40)}$$

The collision frequency ($\nu_c = \gamma_U$) obtainable through Eq. (II.39) from Eq. (II.40) is plotted in Fig. 7. Also indicated in this figure are collision frequencies as obtainable from certain other theories of electron broadening. First let us consider the Unsold result.

We suppose the frequency perturbation due to the

atomic charge distribution at $\mathbf{r}$ to be

$$\Delta\omega = \frac{\epsilon}{\hbar} U(\mathbf{r}).$$

Then the Weisskopf form of the interruption theory (we shall consider this in detail in Chapter IV) yields

$$\gamma_c = N v \pi \rho_0^2 \qquad \text{(II.41)}$$

where ρ_0 is the minimum transit distance for perturbations effective in broadening.

The classical result which has been obtained by Lindholm—and which we shall also consider in Chapter IV—is

$$\gamma_c = N v \int_0^\infty 4\pi\rho d\rho \sin^2\frac{\eta}{2} \qquad \text{(II.42)}$$

and the phase shifts may be calculated from

$$\eta = \frac{2}{v} f(\rho) = \frac{2}{v} \int_0^\infty \{U_{2p}[(\rho^2+x^2)^{\frac{1}{2}}] - U_{1s}[(\rho^2+x^2)^{\frac{1}{2}}]\} dx$$

wherein we remark that particular atomic states have been specified. The potentials are those of the atomic distribution—averaged over angles—that act on the broadening electron specifically

$$U_{2p}(r) = \frac{\epsilon}{a}\left[\frac{1}{r} + \frac{3}{4} + \frac{r}{4} + \frac{r^2}{24}\right] e^{-r}$$

$$U_{1s}(r) = \frac{\epsilon}{a}\left[\frac{1}{r} + 1\right] e^{-2r}.$$

The quantum form of the Lindholm result,

$$\gamma_c = N v \sum_l \sigma_l [1 - \cos(\eta_l - \eta_l')], \qquad \text{(II.43)}$$

also appears in Fig. 7. Meyerott and Margenau have used the calculations of Swan and Massey and Moiseiwitsch in addition to their own for the phase shift evaluations in this equation.

The comparisons between the universal broadening effect of KBM and these other theories should be rather apparent from the figure in question.

Let us now conclude our study of Stark broadening by an inclusion of the second term in Eq. (II.35).

As we have remarked, this term results in what these authors have called "polarization broadening." The line shape is of the same form, and when both the polarization and universal effects are included, one obtains

$$I(\omega) \propto \frac{1}{(\omega-\omega_0)^2 + (\gamma + \gamma_U + \gamma_P)^2}$$

where, as for the universal effect,

$$\gamma = 2 N v \sigma S_P$$

$$S_P \equiv \int_0^{x+} \frac{|G_{nm}|^2}{x^2} dx \qquad \text{(II 44a)}$$

$$G_{nm} = \int e^{i\mathbf{K}\cdot\mathbf{r}} \bar{\psi}_n(r) \psi_m(r) dr. \qquad \text{(II.44b)}$$

KBM have given a comparison of the universal, polarization, and ionic Stark broadening effects for the Lyman α line which we reproduce:

Universal broadening:	$\gamma_U/\gamma = 0.13$
Polarization by reorientation ($2p0 \rightarrow 2p\pm1$):	$\gamma_{P1}/\gamma = 0.027$
Quenching ($2p0 \rightarrow 1s0$):	$\gamma_Q/\gamma \quad 0.0013$
Polarization by induction ($2p0 \rightarrow 2s0$):	$\gamma_{P2}/\gamma = 2.7$
Ionic stark broadening:	$\gamma_{\text{Stark}}/\gamma \doteq 80.$

As concerns γ_{P1}, γ_{P2}, γ_Q we note that they all arise from the second term in Eq. (II.35). The only difference is the specific off-diagonal matrix element which is involved. Depending on what this element is, one can physically qualify the phenomenon involved as "polarization by induction," "polarization by reorientation," or "quenching." For example, if we reorient the dipole from $m=0$ to $m=1$, we obtain the "polarization by reorientation" effect.

III. STATISTICAL BROADENING

A. A First Approach

Any consideration of statistical broadening may perhaps best begin with some mention of the Franck-Condon[9,18] principle, for if we are to point to any one concept as fundamental to this type broadening, then it is to this principle that we must apply ourselves. As one might suppose, the Franck-Condon principle was first enunciated by Franck and enlarged and clarified by Condon. Having startled the reader with this opening statement, let us consider the principle itself.

In doing so we consider the potential curves of Fig. 8. Although the principle in question was first advanced in connection with the constituent atoms in a diatomic

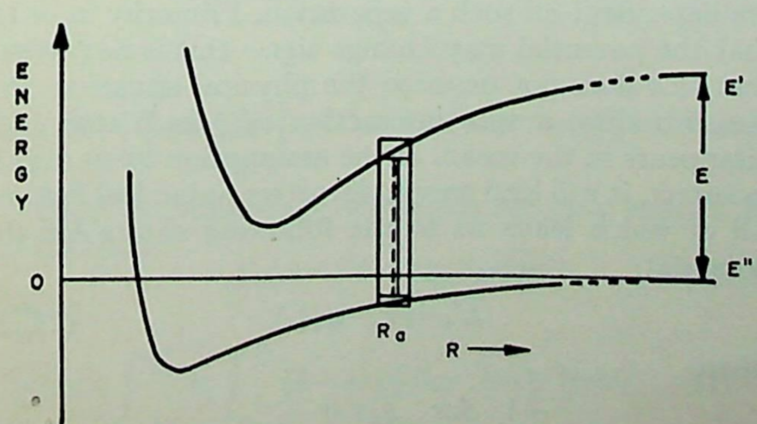


FIG. 8. Potential curves going with two different electronic states for consideration in connection with the statistical theory.

molecule, our considerations will relate to two free atoms. The potential curves of the figure we suppose give the energy, as a function of atomic separation, for two atomic states. In the form in which we shall utilize it, this principle hypothesizes that the atomic transition takes place vertically on the figure, that is, it supposes the interatomic separation R to remain unchanged during this transition.

Jablonski[33] first utilized this principle in advancing a qualitative explanation of spectral line broadening. Now during an electronic transition we suppose (as a consequence of the Franck-Condon principle) that (1) the separation and (2) the relative velocity of the interacting emitter and broadener remain unchanged. Let us again consider Fig. 8 in detailing the fashion in which such a pair of assumptions may be expected to lead to line broadening.

First of all, the level separation at infinite interatomic distance corresponds to the frequency of the emitted radiation in the absence of any broadening perturbations. Next let us suppose the transition to take place when the atomic separation is R_a. Under these conditions the level energy separation will be different from the unperturbed separation, and radiation of a different frequency will be emitted. From this it follows (and we are supposing binary collisions for conceptual clarity) that the line intensity maximum will be shifted to the frequency corresponding to the most probable atomic separation. Around this maximum will then be distributed other frequencies—corresponding to other separations—with the intensity (probability) of these frequencies corresponding to the probabilities for the various separations requisite for the frequencies. As we shall later see detailed the type of spectral line will depend on the type of interaction curve.

This process has been described by Oldenberg[71,72] and Minkowski[69] as one in which the difference between the "natural frequency" of the atom and the frequency observed is made up from the relative velocity of the two particles involved. To our two potential curves—which, we remark, must be different in order that any but one frequency be emitted as a function of separation—there will correspond two different kinetic curves. Now, although the relative velocity will remain constant during the transition, it will change (in that it will be governed by a new kinetic curve) subsequent to the collision. In this fashion one can create a picture of this radiation-kinetic energy exchange.

B. Statistical Theory

In a series of papers on the subject Margenau[55–58,62] has developed the statistical theory from the qualitative explanation of the last section to the sophisticated, contemporary theory.[62] In his earlier studies of the subject, he devoted himself largely to the specific treatment of the van der Waals broadening, primarily, it would seem, because he felt this was the explanation of the broadening by foreign gas atoms possessed of no permanent poles. From our present apparent vantage point, the Holtsmark theory appears to be simply an explanation of Stark broadening which we may neatly categorize and apply at the proper time. At the time it was written, however, it was meant to be a little more than this. It was advanced in order—at least so the author tells us in the original article—to supplant the earlier Michelson-Lorentz theory which had been found wanting in general applications. It was found wanting, as we might suppose, since it requires the possession of a permanent pole on the part of the broadener. Margenau essentially advanced the van der Waals force as a broadening agent quite applicable to such situations and, at the same time, created the theory which was to provide the framework within which this broadening agent could be utilized. Some time later, however, he set forth the development of the generalized theory which we shall detail and from which the earlier work may be obtained.

The broadening we attribute to some sort of forces which act between the two atoms. Now these forces give rise to potential curves of the type which we have discussed in the preceding section. Let us further suppose that the differential energy level distortion due to these forces is given $U=\sum_i U_i$ where U_i is the interaction between the emitter and the ith broadener. The assumption of such additivity of potentials, of course, supposes that there are no interactions between broadeners, an assumption corresponding to the assumption of point broadeners in the simple Holtsmark Stark theory. The mathematics of the statistical theory is now devoted to answering the following question: What is the chance that a configuration of broadeners exists such that a displacement of the spectral line corresponding to U results? This "chance" is, of course, precisely the intensity of the displaced frequency in the broadened line.

In detailing the mathematics of the situation, we must begin by determining an expression for the potential. Firstly, we suppose this potential to depend on some inverse power of the emitter-broadener separation since van der Waals, dipole-dipole, and many other interactions of a type which we shall wish to investigate are dependent on such a separation. Primarily in order that the potential may change sign—this is not to say that this does not describe the physical situation—we also introduce a spin interaction $u(\xi)$ such that $u(\xi)$ disappears in the mean. If the assumption as to $u(\xi)$ is incorrect, it will lead only to incorrect numerical factors. All of which leads us to the following choice for the potential:

$$U_i = a r_i^{-j} u(\xi_i) \tag{III.1}$$

where

$$u(\xi_i) = \begin{matrix} -1 & \text{for} & \xi_i \leqslant 0 \\ +1 & \text{for} & \xi_i > 0 \end{matrix}; \quad -1 \leqslant \xi_i \leqslant 1. \tag{III.2}$$

We suppose the volume occupied by the entire gas to be $(4/3)\pi R^3$. The distribution of broadeners may be taken as random** so that the probability that the emitter-broadener separation be r is

$$\frac{dV}{V}=\frac{r^2 \sin\vartheta d\vartheta d\varphi dr}{(4/3)\pi R^3} \Longleftrightarrow p(r)dr=\frac{3}{R^3}r^2dr \qquad \text{(III.3a)}$$

and of a spin ξ

$$p(\xi)d\xi=\frac{dl}{l}=\tfrac{1}{2}d\xi. \qquad \text{(III.3b)}$$

The probability that the spectral frequency radiated by the emitter differs in energy from the unperturbed frequency by an amount between U and $U+dU$ is

$$I(U)dU=\prod_i \frac{3}{2R^3}\int\int_U r_i^2 dr_i d\xi_i. \qquad \text{(III.4)}$$

The integration in Eq. (III.4) is to be carried out over those portions of space which yield the perturbation in question. This integration may be extended to one over all space by an application of the proper Dirichlet factor

$$I(U)dU=\frac{\zeta}{2R^3}\int\int r_1\cdots r_n dr_1\cdots dr_n d\xi_1\cdots d\xi_n$$
$$\cdot\frac{1}{\pi}\int\frac{\sin[(\tfrac{1}{2}dU)s]}{s}e^{i\Upsilon s}ds$$
$$\doteq\frac{3}{4\pi R^3}dU\int\int\int_{\text{all space}} r_1\cdots r_n e^{i\Upsilon s}dr_1\cdots$$
$$\times dr_n d\xi_1\cdots d\xi_n ds \qquad \text{(III.5a)}$$

where

$$\Upsilon=\sum_i U_i-U. \qquad \text{(III.5b)}$$

Now we have supposed there to be no interactions between broadeners so that the n product integrals here are identical, and we may rewrite Eq. (III.5a) as

$$W_n(U)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-isU}A_n(s)ds \qquad \text{(III.6a)}$$

where

$$A_n(s)=\left\{\int p(\xi)d\xi\int p(r)dr \exp[isar^{-j}u(\xi)]\right\}^n. \qquad \text{(III.6b)}$$

Using no subterfuge whatever, we may rewrite Eq. (III.6b) as

$$A_n(s)=[1-3B(s)/2R^3] \qquad \text{(III.7a)}$$

where

$$B(s)=\int_{-1}^{+1}d\xi\int_0^R \{1-\exp[isar^{-j}u(\xi)]\}r^2dr. \qquad \text{(III.7b)}$$

** A Maxwell-Boltzmann distribution was found by Margenau to have little effect.

Equation (III.7a) may surely be expanded according to the binomial theorem. In the result we allow R to approach infinity which converts our binomial expansion into the infinite MacLaurin series for

$$A(s)=\exp[-2\pi NB(s)] \qquad \text{(III.7c)}$$

so that our Eq. (III.6a) becomes

$$W(U)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp[-isU-2\pi NB(s)]ds. \qquad \text{(III.8)}$$

Next, Eq. (III.7b) may be integrated over ξ so that

$$B(s)=2\int_0^{+\infty}(1-\cos sv)r^2dr \qquad \text{(III.9a)}$$

where

$$v=ar^{-j}. \qquad \text{(III.9b)}$$

If we let

$$sv=\mathfrak{F}, \qquad \text{(III.9c)}$$

B may be written as

$$B=\tfrac{2}{3}(a|s|)^{3/j}\int_0^\infty \mathfrak{F}^{-3/j}\sin\mathfrak{F}d\mathfrak{F} \qquad \text{(III.10)}$$

where integration by parts has been utilized and l'Hospital's rule applied. This application of l'Hospital's rule results in the restriction of j such that $j>\frac{3}{2}$. This is actually no restriction on the theory, since none of our interaction laws will require a value of less than $\frac{3}{2}$.

For convenience, let

$$(4/3)a^{3/j}\int_0^\infty \mathfrak{F}^{-3/j}\sin\mathfrak{F}d\mathfrak{F}=g_j \qquad \text{(III.11a)}$$

so that

$$2\pi B(s)=g_j|s|^{3/j}. \qquad \text{(III.11b)}$$

Our probability $W(U)$ is a real affair, and the utilization of this reality and Eq. (III.11b) in Eq. (III.8) yields

$$W(U)=\frac{1}{\pi}\int_0^\infty \exp[-Ng_js^{3/j}]\cos(sU)ds \qquad \text{(III.12)}$$

since the integrand is an even function.

By evaluating Eq. (III.12) for the case $U=0$ (no shift)

$$W(0)=\frac{j}{3\pi}\Gamma\left(\frac{j}{3}\right)\frac{1}{(Ng_j)^{j/3}}. \qquad \text{(III.13)}$$

Margenau demonstrated the interesting result that the line center intensity decreases as $N^{-j/3}$.

If we define a function I_j as,

$$I_j(x)=\int_0^\infty \exp\left[-\left(\frac{u}{x}\right)^{3/j}\right]\cos su\,du \qquad \text{(III.14a)}$$

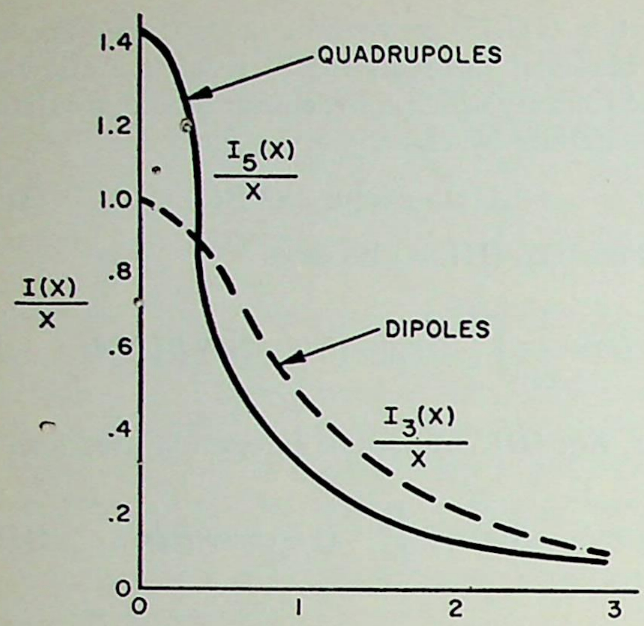


FIG. 9. Dipole and quadrupole line shapes according to the statistical theory. (After Margenau.[62])

then Eq. (III.12) may quite readily be written as

$$W(U)=\frac{1}{\pi|U|}I_j\left(\frac{U}{U_0}\right) \qquad \text{(III.14b)}$$

where

$$U_0=(Ng_j)^{j/3}. \qquad \text{(III.14c)}$$

All of which leads us to the departure point of special cases, and, with Margenau, we shall consider three, namely, broadening by (1) permanent dipoles, (2) quadrupoles, and (3) forces which do not change sign.

In the dipole case we have $j=3$ so that from Eqs. (III.14c), (III.11), and (III.14a)

$$U_0=\tfrac{2}{3}\pi^2Na; \quad I_3(x)=\frac{x}{1+x^2}$$

so that Eq. (III.14b) becomes

$$W(U)=\frac{U_0}{\pi(U^2+U_0^2)} \qquad \text{(III.15a)}$$

or, when we recall that $U=\hbar(\omega_0-\omega)$

$$W(\omega)=\frac{U_0/\hbar^2\pi}{(\omega_0-\omega)^2+(U_0/\hbar)^2}. \qquad \text{(III.15b)}$$

This is of some interest, since it is the Michelson-Lorentz or dispersion form of the line shape.

A consideration of Eq. (III.11a) and Eq. (III.14a) should suffice to justify the statement that numerical calculations or series expansions are requisite for the evaluation of the integrals in these equations for other values of j. Margenau has carried out this numerical evaluation for the quadrupole case, $j=5$, and the resulting curve $I_5(x)/x$ is compared to the one for the dipole case in Fig. 9.

As our third example we consider the van der Waals force as the broadening agent. In this instance we take the same sign so that the spin interaction factor in the potential energy expression is simply equated to unity. As we have remarked the statistical theory was originally developed by Margenau in conjunction with a consideration of these van der Waals forces.

Under these conditions we obtain in place of Eq. (III.10)

$$B(s)=2\int_0^\infty (1-e^{isv})r^2dr$$
$$=-i\tfrac{2}{3}(a|s|)^{3/i}\int_0^\infty \mathfrak{F}^{-3/i}\exp(i\mathfrak{F})d\mathfrak{F}$$
$$=\tfrac{1}{2}|s|^{3/i}g=\frac{1}{2\pi}|s|^{3/i}(g'+ig'') \qquad \text{(III.16)}$$

corresponding to which one obtains

$$W(U)=\frac{1}{\pi}\int_0^\infty \exp(-Ng's^{3/i})$$
$$\times\cos(sU+g''s^{3/i}N)ds \qquad \text{(III.17)}$$

for forces which do not change sign. For the van der Waals interaction $m=6$, and Eq. (III.17) may be evaluated utilizing the g' and g'' values obtainable from Eq. (III.16) with the result

$$I(\Delta\nu)=\frac{1}{\pi}\int_0^\infty \exp[-\tfrac{2}{3}\pi N(2\pi as)^{\frac{1}{2}}]$$
$$\times\cos[s(\Delta\nu)-\tfrac{2}{3}\pi N(2nas)^{\frac{1}{2}}]ds$$
$$=\frac{2\pi\sqrt{a}}{(\nu-\nu_0)^{\frac{3}{2}}}N\exp\left[-\frac{4}{9}\frac{\pi^3aN^2}{(\nu-\nu_0)}\right] \qquad \text{(III.18)}$$

which is precisely the result Margenau had originally obtained for this special case. The half-width of this distribution is

$$\delta=0.82\pi^3aN^2 \qquad \text{(III.19a)}$$

and the line shift

$$\Delta\nu_{\max}=(\tfrac{2}{3}\pi)^3aN^2. \qquad \text{(III.19b)}$$

We have compared the van der Waals distribution predicted by the theory with an observed line in Fig. 10. Although the good agreement in the red wing is ap-

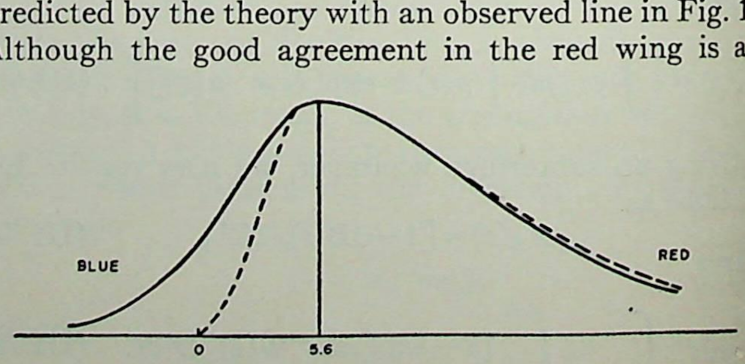


FIG. 10. A comparison of the theoretical prediction of the statistical theory for the van der Waals forces with an experimental result. (After Margenau.[56])

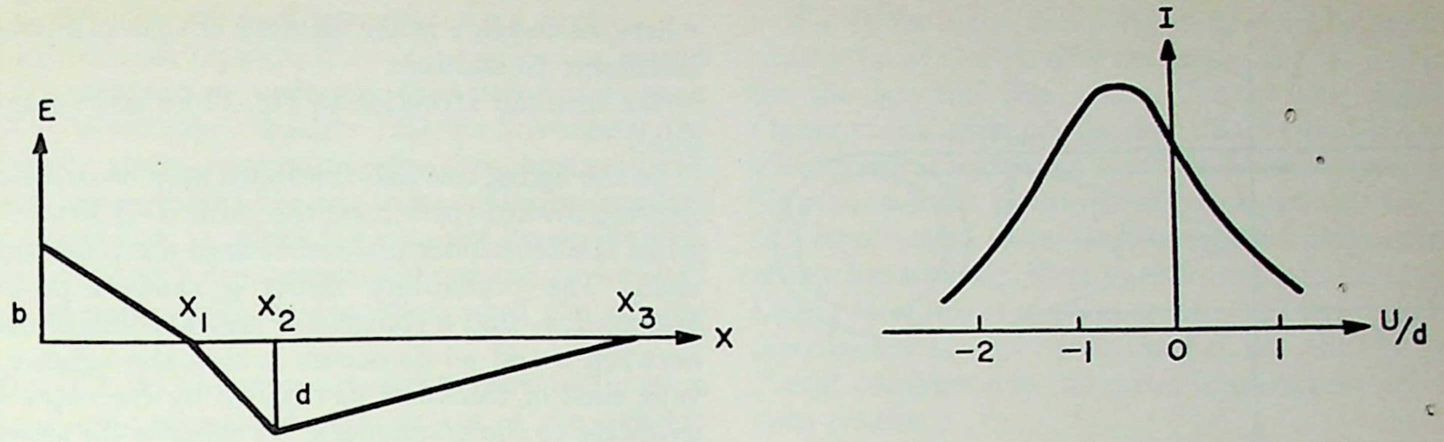


FIG. 11. An example of Margenau's application of the integrable block and triangle function to the statistical theory. (After Margenau.[57])

parent, we note the poor agreement between theory and experiment in the blue wing. Albeit other limits (such as the low velocity approximation) are imposed on the theory, one reason for the discrepancies appears worthy of mention at this point. Were the theory itself completely free of limiting assumptions, the precise potential curve for a given broadening situation would have to be known in order that a precise agreement between the theoretical results and the observations could be obtained. In general (although one could probably say in all cases) the potential curve due to the interaction between emitter and surrounding broadeners is known with but little accuracy. Thus, an approximate curve must be used which may be a reasonable assumption but which cannot be expected to detail the behavior of the potential energy and from which we can hardly expect to obtain completely correct broadening answers. Along this line, Margenau had, in one of his early studies of the situation, considered a block and triangle approximation to the interaction potential curves and then numerically integrated the results of inserting these approximations into the line shape integrals. As an interesting example of the method we give a few of his results in Fig. 11. Particularly in the study of forces by means of line shapes does it appear that this technique might have some interesting possibilities.

C. The Jablonski Theory

Jablonski felt that a wave mechanical theory of line broadening should be developed and that this development might well follow the line of attack utilized by certain authors in their treatment of molecular rotation-vibration spectra. To this end Jablonski's paper of 1937[34] and those which followed[35–38] were devoted.††

We begin, with Jablonski, by assuming that our absorber (emitter) and the N' foreign gas atoms, whose perturbing influence shall be responsible for the broadening of the spectral line, go to make up a very large, $(N'+1)$-atomic molecule.

†† One of these articles[36] was not available to the author. The material contained in this article, however, is covered by subsequent papers which were consulted.

In treating the molecule Jablonski applied the technique which had originally been developed in connection with molecular electronic-vibration theory, that is to say, he supposed the $(N'+1)$-atomic molecular eigenfunction to be a product of one electronic and N' two atomic eigenfunctions each of these latter is concerned with the interaction between the emitter and one of its perturbers. Since such is the case, the probability for the shift in the emitted radiation frequency will be partially dependent on the matrix element whose evaluation is possible through a utilization of these eigenfunctions. The calculation inferred is one familiar from inelastic scattering theory.

We now suppose that we know the probability distributions $P_1(x_1)$, $P_2(x_2)$, $\cdots$, $P_N(x_{N'})$ of the quantities $x_1, x_2, \cdots, x_{N'}$. Then the problem with which Jablonski was faced was the determination of the probability distribution of

$$X=\sum_{i=1}^{N'} x_i$$

which by the judicious utilization of an inductive proof he was able to show was

$$\int_{\Delta X} P(X)dX=\int_{\Delta X}\left[\delta(X)e^{-\zeta N'}+\sum_{m=1}^{N'}\binom{N'}{m}\times W^{(m)}(X)e^{-\zeta(N'-m)}\right]dX. \quad \text{(III.20)}$$

Let us make the connection between Eq. (III.20) and the development somewhat more apparent. To begin with we shall wish that $x_i=E_i'-E_i$, that is, this variable refers to the translational energy separation between two levels à propos to the two particle system consisting of the emitter and the ith broadener. $\int_{\Delta X} P(X)dX$ gives the probability for the small indicated range of the variable for the translational energy of the $N'+1$ particles comprising the emitting-broadening system. As to the terms within the right-hand integral the first term represents the probability for no translational energy change in the system; the first

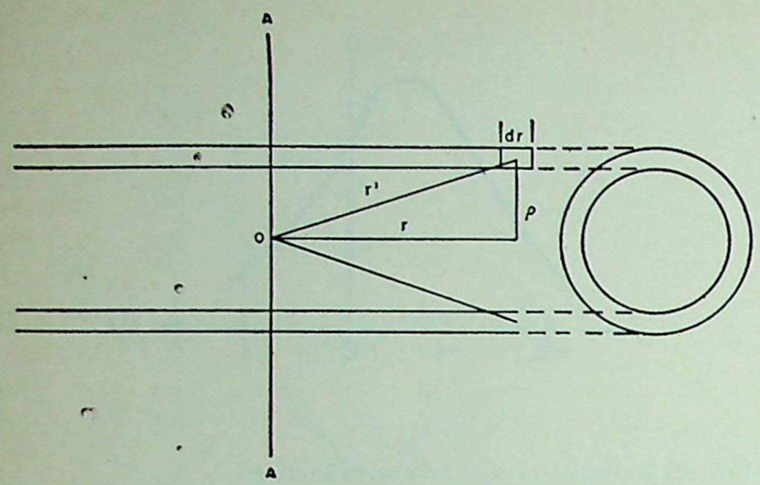


FIG. 12. A physical model for the calculation of the optical collision diameter probability.

term within the sum (the second term on the right of the equation) gives the probability of a binary collision, that is, of a collision between the emitter and one broadener; the second term within the sum gives the probability for an emitter-two-broadener collision, and so on. The probability distribution in question then allocates the requisite relative importance to the various types of collisions and may certainly be of some interest under certain broadening situations, although it has not been utilized in any problems. What we now do is effectively ignore our result, Eq. (III.20), and again suppose binary collisions. Thus, we content ourselves with an interest in $W^{(1)}(X)$.

Now the probability that the emitter-broadener system in a state $n'l$ shall undergo a transition to a state $n''l$ is given by

$$W^{(1)}(X)=W^{(1)}(E_{n'}-E_{n''})$$
$$=\sum_{l=0}^{l_{\max}} Q(l)\frac{\mathfrak{D}_{n'n''}{}^2}{\sum_{n'=1}^{\infty}\mathfrak{D}_{n'n''}{}^2}\frac{dn'}{dE_{n'}}$$
$$=\int_0^{l_{\max}} Q(l)\frac{\mathfrak{D}_{n'n''}{}^2}{S}\frac{dn'}{dE_{n'}}dl$$
$$=\int_0^{l_{\max}} Q(l)A_{n'n''}{}^2\frac{dn'}{dE_{n'}}dl \quad \text{(III.21)}$$

wherein we have utilized the relations $S=\mathfrak{M}_0{}^2$, $\mathfrak{D}_{n'n''}\doteq\mathfrak{M}_0A_{n'n''}$, and $\mathfrak{D}_{n'n''}$ is the matrix element for the electronic transition. $Q(l)$ is the probability of occurrence of a certain l and $dn'/dE_{n'}$ the density of levels in the upper state. In the case of emission we replace $dn'/dE_{n'}$ by $dn''/dE_{n''}$.

We consider probability $Q(l)$ of a particular l value. Firstly, let us make an appeal to correspondences (on the assumption of high l values) in order to equate the classical and quantum angular momenta:

$$(2mE_i)^{1/2}\rho=[l(l+1)]^{1/2}\hbar \Longleftrightarrow Q(l)dl\doteq Q'(\rho)d\rho \quad \text{(III.22)}$$

where, as usual, ρ is the distance of closest approach of broadener to emitter.

We turn our attention to Fig. 12 for an evaluation of $Q(\rho)$.

In the figure, one half the atoms may be considered as moving toward (or across) the plane A–A from the left, while the remainder proceed toward the plane from the right. The probability under a random probability assumption, that a collision of optical collision diameter between ρ and $\rho+d\rho$ occurs is thus the volume of the tube shell of thickness $d\rho$ divided by the total volume available to the broadeners. We suppose the assemblage to be confined to a sphere of radius R. In consequence

$$Q'(\rho)d\rho=\frac{\int_0^{r_{\max}}2\pi\rho d\rho dr}{(4/3)\pi R^3}=\frac{3}{R^3}r_{\max}\rho d\rho$$

from which when we suppose

$$Q'(\rho)d\rho=\frac{3}{2R^2}d(\rho^2)\Longleftrightarrow Q'(\rho)d\rho$$
$$=\frac{3\hbar^2}{4R^2mE}(2l+1)dl. \quad \text{(III.23)}$$

Next the evaluation of $A_{n'n''}$ leads, for Eq. (III.21), to the result

$$W^{(1)}(X)=\frac{3r_c{}^2}{R^3\left|\frac{dX}{dr}\right|_{r=r_c}}\left(1-\frac{U''(r_c)}{E_{n''}}\right)^{1/2}$$
$$\times\left[1-\left(1-\frac{l_1(l_1+1)}{l_t(l_t+1)}\right)^{1/2}\right]+\zeta(X,l_t). \quad \text{(III.24)}$$

The asymptotic form $(l=l_t)$ of $W^{(1)}(X)$ is

$$W^{(1)}(X)=\frac{3r_c{}^2(1-U''(r_c)/E_{n''})^{1/2}}{R^3\left|\frac{dX}{dr}\right|_{r=r_c}} \quad \text{(III.25)}$$

and a corresponding form results for the emission line.

Although it is from Eq. (III.25) that we shall obtain the line shape, let us remark the implications of a complete calculation.

It has been tacitly assumed that there is one interaction curve for the upper electronic state and one for the lower. This is not generally the case for there may well be several interaction curves for each electronic state, and for single encounters we would then have various $W_i{}^{(1)}(X)$ (see Eq. (III.20) for the various pairs of curves. Our $W^{(1)}(X)$ would then have the form

$$W^{(1)}=\sum a_iW_i{}^{(1)}(X). \quad \text{(III.26)}$$

It should also be apparent that since each $A_{n'n''}$ would be different for each $W_i^{(1)}(X)$, each $A_{n'n''}$ would have to comprise a separate calculation for each $W_i^{(1)}(X)$. When one considers that we have here discussed only single encounters, the complexity of the accurate computation for multiple collisions appears rather staggering. In general, we do not know one such pair of curves with any accuracy, however, so we shall content ourselves with having mentioned the general case.

In Eq. (III.25) we now suppose the potential function to depend only on some inverse power of the collision partner separation—$\Delta\omega = (1/\hbar)X = K/r_c^\eta$—so that this equation becomes

$$I(\omega) = \frac{4\pi N K^{3/\eta}}{\eta(\Delta\omega)^{\eta+3/\eta}}\left(1+\frac{2a''\Delta\omega}{3KkT}\right)^{\frac{1}{2}}. \qquad \text{(III.27)}$$

Equation (III.27) corresponds to the statistical result after an approximate fashion in the wings of the line, the second circumstance restricting the theory to this region.

The fact that the Jablonski theory simply leads via another route to the Margenau statistical results renders its practical value somewhat limited. However, the worth of the theory appears to lie in the quantum treatment involved, and, perhaps, some additional results may be drawn from the theory itself by a slightly different use of the approximations or the replacement of certain of them. Foley used certain approximations to show that his form of the Lorentz result could be obtained from the Jablonski theory. This led to a mild controversy, but the flaw in the theory which we have just considered lies not so much in the fact that a judicious utilization of approximations results in someone else's theory—this is usually the case—but in the necessity for the approximations which must be used to obtain any results whatever, for these are the limiting features.

IV. INTERRUPTION BROADENING

A. The Interruption Theory for Zero Collision Time

Now we have remarked in Chapter III that the Holtsmark Stark broadening theory was originally advanced partly due to the belief on Holtsmark's part that the Lorentz—and consequently the Michelson—theories could not be expected to withstand the advent of the quantum theory. Beginning with Lenz's appeal to correspondences in 1924,[46] however, a large number of authors have devoted themselves to demonstrating the incorrectness of such a viewpoint, quite often in no uncertain terms. We shall give one short consideration of quantum classical equivalence a bit later, but now we consider the early work of Lenz only in order to extract from it his major contribution to the development of the interruption theory.

We have seen the fashion in which Michelson's utilization of the Fourier transform results in Eq. (I.8) for the spectral line shape. Let us also remark the Lorentz utilization of a random distribution of electronic vibrational coordinate and time derivative thereof. Now our further interruption considerations will consist to a large extent in developing the Michelson conception of line broadening. First in this development then is the Lenz procedure of averaging Eq. (I.8) over the Lorentz distribution of τ. If we let the I of Eq. (I.8) be given by I' and consider τ as the mean intercollision time, there then results

$$I(\nu) = \int_0^\infty I'(\nu)e^{-\tau'/\tau}d\tau' = \text{const}\frac{(1/2\pi\tau)}{(\nu-\nu_0)^2+(1/2\pi\tau)^2} \qquad \text{(IV.1)}$$

wherein the exponential function defines the Lorentz distribution.

This rather familiar line shape equation has as its half-width

$$\delta = \frac{1}{\pi\tau} = \frac{1}{\pi}\frac{\langle v\rangle}{l} = \langle v\rangle N\rho^2 \qquad \text{(IV.2)}$$

wherein $\langle v\rangle$ is the relative velocity. The symbol l refers to the mean free path. In defining the symbol ρ we shall find the next refinement in the theory arising quite naturally and almost unbidden.

In the Michelson consideration we recall that ρ is simply the atomic diameter since the atom was considered a billiard ball sort of thing whose diameter occurs specifically in the expression for the free path. This is admittedly a more strained physical interpretation than the one subsequently given by Weisskopf.[92,93]

Now surely we can expect, say, the van der Waals forces between an emitter and a broadener to exert some influence on the broadening of a spectral line before the two collision partners are separated by an atomic diameter, and we must suppose ourselves capable of defining this latter quantity. Weisskopf suggested that we define ρ as the "optical collision diameter," a quantity which defines an atomic collision for purposes of line broadening and which we now proceed to consider in some detail.

We begin by considering the quantity ω_0 in Eq. (I.6b). This quantity, of course, represents the angular vibrational frequency of the spectrum producing valence electron. In the Michelson treatment it was a constant, since nothing was present which could change it until the sudden bang of the billiard ball collision. With the Weisskopf assumption of action at a distance through some sort of interaction force, however, this situation is changed. As a result of this force ω_0 will begin to change with time (this temporal coordinate is introduced by the heat motion of the emitter) as soon as the emitter is within the range of these forces. Now if these forces are,

say, a function of emitter broadener separation—which we generally consider them—of the form $\Delta(r)$ then

$$\omega_0=\omega_0'+\int_0^t \Delta(r)dt'=\omega_0'+\eta \qquad \text{(IV.3)}$$

wherein ω_0' is the vibrational frequency of the photoelectron in the absence of any perturbing forces while the remaining term is the change in the phase of this vibration due to perturbing influence of the collision partners. Perhaps redundantly, let us note that for conceptual clarity we are considering binary collisions. All of which means that the phase of the emitted radiation will change with the approach of the emitter. And at this point we introduce the crux of the Weisskopf argument: A collision has been undergone by the emitter when the phase of the emitted radiation, η, has changed by unity. This is an admittedly arbitrary collision definition, but one which has proven quite satisfactory for certain physical situations and, further, one which forms one of the bases for the simple form of the interruption theory.

Having defined an optical collision we now simply consider the ω_0 of Eq. (I.6b) again a constant between two such collisions in order to obtain once more Eqs. (IV.1) and (IV.2) for the line shape and half-width. The optical collision diameter which we shall now evaluate from our definition of an optical collision has been changed by this new collision definition. As an example, let us determine the optical collision diameter for use in Eq. (IV.2) for the case of van der Waals forces. For this case $\Delta(r)=C/r^6$ in Eq. (IV.3). If we now let $x=\langle v\rangle t/\rho$, we may then write

$$\eta=\int_{-\infty}^{+\infty}\frac{C\,dt}{(\langle v\rangle^2t^2+\rho^2)^3}=\frac{C}{\rho^5\langle v\rangle}\int_{-\infty}^{+\infty}\frac{dx}{(x^2+1)^3}$$

$$=\frac{C}{\rho^5\langle v\rangle}\left(\frac{3\pi}{8}\right)\sim 1$$

and wherein we see that from our definition of a collision, the closest approach occurs at $t=0$, from which

$$\rho=\left(\frac{3\pi C}{8\langle v\rangle}\right)^{1/5}$$

and we remark that C is in angular frequency units, ω.

In this section then we have discussed the Lenz-Weisskopf modifications of the Michelson theory which resulted in what may be called the simple form of the interruption theory. This simple theory certainly has its areas of application which may roughly be described as low pressures and frequencies near line center. Now the reasons for such restrictions to the theory may be rendered quite physically plausible by a reference to Fig. 1.

In our ruminations of Chapter I we considered points "b" and "c" as positions at which billiard ball collisions occurred. Now we simply consider these locations as defining the optical collisions of this section. The remainder of our earlier description of the figure is precisely the same. Now then, it is apparent that such a physical picture of the optical collision process is possessed of the following flaws: (1) It neglects the effects of distant collisions. In this connection let us consider the "A" broadeners. They will actually have some effect on the resulting spectral line, but, since they are too far removed from the emitter path to cause phase shifts as great as unity, their presence is neglected by the theory. (2) The effect of close collisions is also neglected. At point "c" the emitter is at such a distance from the "B" broadener that the phase has changed by unity. As a result we declare a collision to have occurred and cut off the wave train. Thus, we neglect the effect of greater phase changes. This second neglect is essentially the same as neglecting the time of collision. A further consideration of the physical situation suffices to demonstrate that (1) limits our application of the theory to relatively low pressures while (2) restricts the application in the line wing. This second restriction arises since surely we may expect the large frequency perturbations of the line wing to arise as a result of phase shifts greater than unity. Although, as we have remarked, there is a definite region of applicability of this theory, there are refinements which result in theories of more comprehensive application. The first and perhaps most obvious of these is the inclusion of the time of collision. Such an inclusion was first attempted by Lenz[47] in an effort to obtain an interruption theory which would allow for the shifts and asymmetries often observed in spectral lines and totally unprovided for in this simple theory. He did, indeed, achieve his end, but, unfortunately at the expense of applicability, for his result was even more restricted than those which we have already considered. Lindholm[50] showed that these shifts and asymmetries may be obtained from the interruption theory without considering the time of collision. Actually, he eliminated approximation (1) rather than approximation (2) as we shall see.

B. The Inclusion of Various Phase Shifts

Let us begin by rewriting Eq. (I.6b) with the help of Eq. (IV.3) as

$$J(\omega)=\int_{-\infty}^{+\infty} e^{i2\pi(\nu-\nu_0)t+i\eta(t)}dt$$

from which we may obtain for the spectral line shape:

$$I(\nu)=|J(\omega)|^2$$

$$=\iint_{-\infty}^{+\infty} e^{i2\pi(\nu-\nu_0)(t''-t')+i[\eta(t'')-\eta(t')]}dt'dt''$$

$$=\int_{-\infty}^{+\infty} e^{i2\pi(\nu-\nu_0)t}dt\int_{-\infty}^{+\infty} e^{i[\eta(t+t')-\eta(t')]}dt'. \qquad \text{(IV.4)}$$

We now consider the integral over dt' in Eq. (IV.4). It was Lenz who developed the ingenious method of evaluation which we shall presently detail: Lindholm later utilized it, and Foley[17] appears to have named it. At any rate the Lenz idea may be described as follows.

If we suppose a random distribution in time, then the integral in t' has the look of a time average, specifically, the time average of the exponential. Now let us recall Boltzmann's ergodic hypothesis. A consideration of this hypothesis in connection with the average in question suffices to demonstrate the equality of this time average and a statistical average of the exponential. Quite simply then we evaluate this integral by determining the statistical average of the exponential. We will surely agree it legitimate to equate the average to $\exp[-A(t)+iB(t)]$. Since $I(\nu)$ and $\eta(t)$ are real, we require $A(-t)=A(t)$ and $B(-t)=-B(t)$ so that

$$I(\nu)=\int_0^\infty e^{-A(t)}\cos[2\pi(\nu-\nu_0)t+B(t)]dt. \quad \text{(IV.5)}$$

Next it is assumed that but three different phase changes occur on collision.‡‡ These we shall designate as η_a, η_b, η_c. This in turn would mean that three different "differential collision cross sections," σ_i occur.

The total cross section is $\sigma=\sigma_a+\sigma_b+\sigma_c$, and the mean time between collisions is τ. The probability that n collisions of type a corresponding to a phase shift η_a (a collisions), m b collisions, and l c collisions will occur in time t is

$$\left(\frac{\sigma_a}{\sigma}\right)^n\left(\frac{\sigma_b}{\sigma}\right)^m\left(\frac{\sigma_c}{\sigma}\right)^l\frac{1}{n!m!l!}\left(\frac{t}{\tau}\right)^{n+m+l}e^{-t/\tau}. \quad \text{(IV.6)}$$

Further, our exponential may be evaluated as

$$\eta(t+t')-\eta(t')=\sum_k[\eta_k(t+t')-\eta_k(t')]$$
$$=n\eta_a+m\eta_b+l\eta_c. \quad \text{(IV.7)}$$

In this case then the mean value of $e^{i[\eta(t+t')-\eta(t')]}$ which Lindholm assumed equal to the second integral in Eq. (IV.4) is given by

$$\langle e^{i[\eta(t+t')-\eta(t')]}\rangle$$
$$=\sum_{n=0}^\infty\sum_{m=0}^\infty\sum_{l=0}^\infty\left(\frac{\sigma_a}{\sigma}\right)^n\left(\frac{\sigma_b}{\sigma}\right)^m\left(\frac{\sigma_c}{\sigma}\right)^l$$
$$\times\frac{1}{n!m!l!}\left(\frac{t}{\tau}\right)^{n+m+l}e^{-t/\tau}\{e^{i[n\eta_a+m\eta_b+l\eta_c]}\}. \quad \text{(IV.8)}$$

Since

$$\sum_{n=0}^\infty\left(\frac{\sigma_a t e^{i\eta_a}}{\sigma\tau}\right)\frac{1}{n!}=\exp\left[\frac{\sigma_a t}{\sigma\tau}e^{i\eta_a}\right], \text{ etc.}$$

‡‡ As we shall see, we could initially assume some other number and later extend the number of allowed phase changes as desired.

Equation (IV.8) becomes

$$\langle e^{i[\eta(t+t')-\eta(t')]}\rangle$$
$$=\exp\left\{\frac{t}{\tau\sigma}[\sigma_a e^{i\eta_a}+\sigma_b e^{i\eta_b}+\sigma_c e^{i\eta_c}]\right\}. \quad \text{(IV.9)}$$

Utilizing Eq. (IV.9) in Eq. (IV.4) we obtain

$$I(\nu)=\Re\int e^{2\pi i(\nu-\nu_0)}$$
$$\times\exp\left\{\frac{t}{\tau\sigma}[\sum_i\sigma_i(\cos\eta_i+i\sin\eta_i)-\sigma]\right\}dt$$
$$=\int\exp\left\{-\frac{t}{\tau\sigma}\sum_i[\sigma-\sigma_i\cos\eta_i]\right\}$$
$$\times\cos\left[\frac{t}{\tau\sigma}\sum_i\sigma_i\sin\eta_i+2\pi(\nu-\nu_0)\right]dt \quad \text{(IV.10)}$$

where we have taken the real part of $\exp(i\sin\eta_i)$ since the intensity is a real quantity. Thus, from Eqs. (IV.5) and (IV.10) there results

$$A=t\sum_i\frac{\sigma_i}{\sigma\tau}[\sigma-\sigma_i\cos\eta_i]=\alpha t \quad \text{(IV.11a)}$$

$$B=t\sum_i\frac{\sigma_i}{\sigma\tau}\sin\eta_i=\beta t, \quad \text{(IV.11b)}$$

and Eq. (IV.5) becomes

$$I(\nu)=\int_0^\infty e^{-\alpha t}\cos\{[2\pi(\nu-\nu_0)+\beta]t\}dt$$
$$=\frac{\text{const}}{\left[(\nu-\nu_0)+\frac{\beta}{2\pi}\right]^2+\left[\frac{\alpha}{2\pi}\right]^2}. \quad \text{(IV.12)}$$

In obtaining Eq. (IV.12) the collision time has been neglected in that we simply took the various phase changes as having occurred, thus neglecting the effects of close collisions. Equation (IV.12) of course, yields no line asymmetry. This we shall obtain on consideration of close collisions.

It might also be noted that we restricted ourselves to three phase changes η_a, η_b, and η_c, in setting up Eq. (IV.8), but there is no reason to restrict Eqs. (IV.11) for the shift and half-width in this manner.

In Sec. A of this chapter we have remarked the fashion in which the simple interruption theory ignores the effects of (1) distant and (2) close collisions. In the developement which we have just considered the effect of this ignoration of distant collisions has been demonstrated. In other words we have demonstrated the

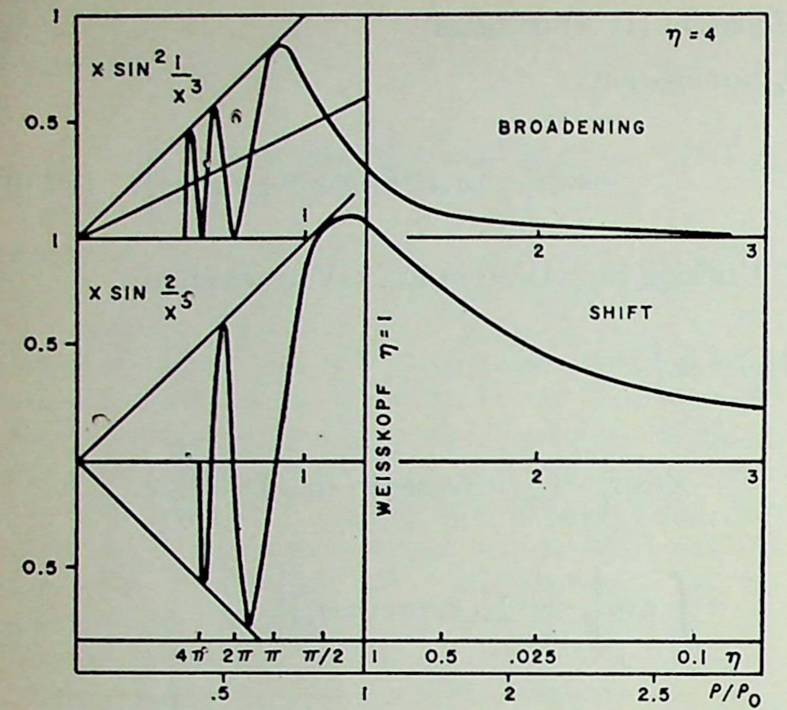


FIG. 13. An illustration of the reason for the failure of the simple interruption theory to yield a line shift. (After Lindholm[99] and Unsold.[97])

fashion in which consideration of these distant collisions leads to a line shift. This fact is further borne out by a study of Fig. 13.

The vertical line in about the center of the figure is the Weisskopf collision diameter. Now the area under the upper curve is related to the integral in Eq. (IV.11a) and is hence a measure of the broadening. In like manner the area under the lower curve is a measure of the shift. The simple theory only considers those areas to the left of the abscissa specifying the Weisskopf collision diameter. It thus becomes quite apparent that with the adoption of this diameter we obtain almost all the broadening effects (insofar as this theory is concerned) while we obtain almost none of the shift effects. In addition these curves serve to illustrate the relative importance of near and distant collisions in broadening and shift.

C. The Effect of Close Collisions

Next Lindholm[51] considered the problem with the collision time included, thus essentially including the effects of both near and distant collisions.

Although our consideration of this more general interruption treatment shall be quite sketchy, we particularly want to emphasize Lindholm's manner of including the time of collision (close collisions) within the framework of his earlier theory.

Let the collision time be t_i, and the phase change per unit of time during the collision be k_i. Then the total phase change in collision will be $\eta_i = k_i t_i$. Let us consider Fig. 14. The coordinates x and y measure time back from t' and t'' respectively as shown. The five arrows below the temporal axis represent five different collisions whose duration is represented by the respective arrow lengths. Only those collisions which occur at least partly during $t''-t'$ are to be included.

It can be seen directly from our definitions that these collisions will contribute to $e^{i[\eta(t'')-\eta(t')]}$ the factor

$$e^{i n_x i k_i \varphi(t_i - x)} \tag{IV.13a}$$

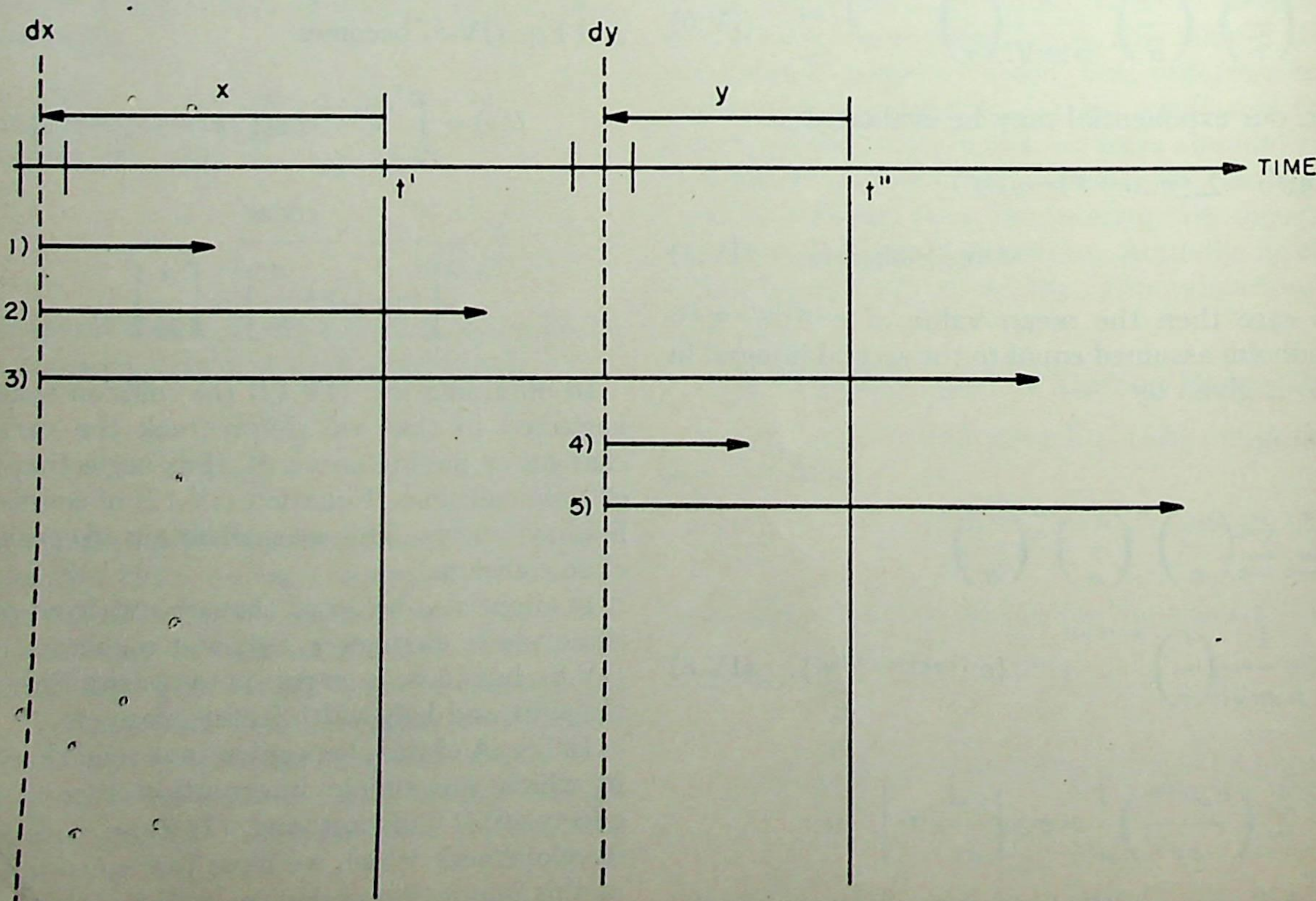


FIG. 14. The Lindholm method of classifying collisions as to duration. (After Lindholm.[51])

where

$$\varphi(t_i-x)=0 \quad \text{for} \quad t_i-x<0$$
$$\varphi(t_i-x)=t_i-x \quad \text{for} \quad 0<t_i-x<t''-t' \qquad \text{(IV.13b)}$$
$$\varphi(t_i-x)=t \quad \text{for} \quad t''-t'<t_i-x.$$

Equation (IV.13b) is merely a restatement of our restriction of the collision such that at least a portion of it occurs within $t''-t'$, and the method of including close collisions has been demonstrated.

In analogy to Eq. (IV.8) one obtains for the present case

$$\langle e^{i[\eta(t'')-\eta(t')]}\rangle$$
$$=\sum_{n_{x(1)1}=0}^{\infty}\cdots\sum_{n_{x(1)i}=0}^{\infty}\sum_{n_{x(2)i}=0}^{\infty}\cdots\sum_{n_{y(j)i}=0}^{\infty}\prod_{i}\prod_{dx(l)}\prod_{dy(j)}$$
$$\times\frac{1}{n_{x(l)i}!e^{dx(l)/\tau}}\left(\frac{\sigma_i}{\sigma}\frac{dx(l)}{\tau}\right)^{n_{x(l)i}}$$
$$\cdot\frac{1}{n_{y(j)i}e^{dy(i)/\tau}}\left(\frac{\sigma_i}{\sigma}\frac{dy(j)}{\tau}\right)^{n_{y(j)i}}$$
$$\times e^{i[n_{x(l)i}k_i\varphi(t_i-x)+n_{y(j)i}k_i\psi(t_i-y)]}. \qquad \text{(IV.14)}$$

In Eq. (IV.14) the $x(l)$ represent the possible dx time intervals during which included collisions may originate. The $y(j)$ represent the possible dy time intervals during which included collisions may originate. Thus, $n_{x(l)1}$ is the number of collisions of the type one originating with some probability during the $dx(l)$ time interval.

Equation (IV.14) in conjunction with Eqs. (IV.4) and (IV.5) may be evaluated for the van der Waals case. For high pressures the result is

$$I(\nu)=I_0=c_1N\langle v\rangle^{6/5}(\Delta\nu)^{-3/2}\exp(-N^2c_2/\Delta\nu). \qquad \text{(IV.15)}$$

A consideration of Eq. (IV.15) shows that Lindholm's line shape has reduced to Margenau's statistical shape for the case of high pressure. If we let the phase shift $\mathfrak{H}=16/3\pi=1.7$ instead of the 0.75 which Lenz had used, we obtain exact agreement between the exponentials in the Margenau statistical theory Eq. (IV.15). Since the statistical theory had yielded such excellent agreement with experiment at high pressure Lindholm chose $16/3\pi$ as the value for $\mathfrak{H}$ on the basis of this resulting agreement. Lindholm justified his disagreement as follows.

First, let us recall that the time duration of the collision is essentially $2\mathfrak{H}$ and consider Fig. 15. Curve (a) in this figure is the actual curve for the phase integral $2\pi b\int dt/(\langle v\rangle^2t^2+\rho^2)^3$. Curve (b) is Lenz's approximate curve where $\mathfrak{H}=0.75$, and curve (c) is the Lindholm curve where $\mathfrak{H}=1.7$. Lindholm felt that the larger $\mathfrak{H}$ value is more justified in that it includes the slow frequency shifts at either end of the true curve (a).

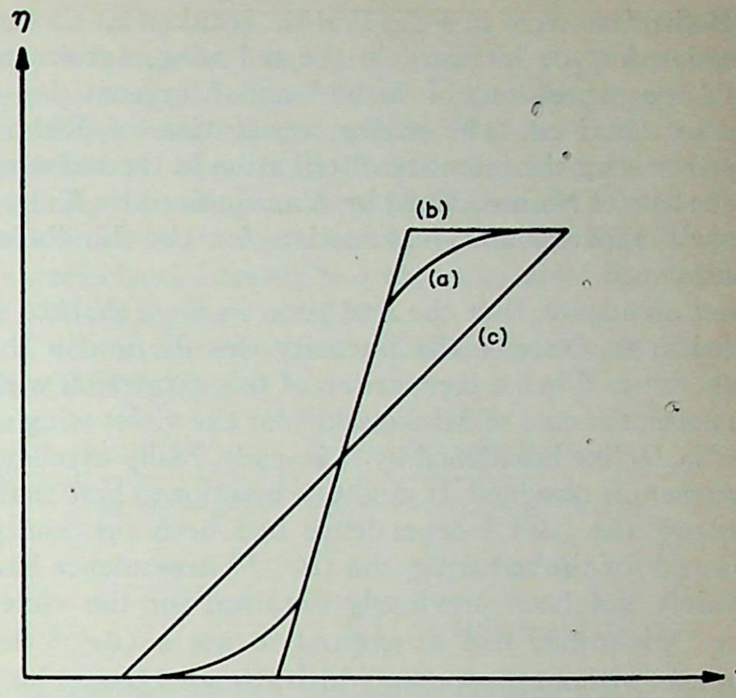


FIG. 15. The Lenz and Lindholm approximations to the phase shift function. (After Lindholm.[51])

The low pressure solution for van der Waals forces yielded

$$I_\nu=I_0+I_\infty$$
$$=0.795\frac{1-(k+0.577l)\dfrac{0.227}{0.795}-0.165l}{(0.795l)^2+(0.577l+k)^2}, \qquad \text{(IV.16)}$$

where the definitions of l and k may be inferred from Eqs. (IV.17). The line shift and half-width are given by

$$\Delta\nu_0=\frac{0.577}{\dfrac{2\pi\mathfrak{H}}{\langle v\rangle}\left[\dfrac{3\pi^2b}{8\langle v\rangle}\right]^{1/5}}=\frac{0.577l}{c}=0.97N\langle v\rangle^{3/5}b^{2/5} \qquad \text{(IV.17a)}$$

$$\tfrac{1}{2}\delta=\frac{0.795l}{c}=2.68N\langle v\rangle^{3/5}b^{2/5} \qquad \text{(IV.17b)}$$

which is essentially the Lenz result.

Having considered the Lenz and Margenau limiting cases, let us turn our attention to the behavior of the van der Waals result in the wings of the line.

For this case, Lindholm was able to use an approximate numerical evaluation to obtain the following two equations for the red and violet wings of the line, respectively,

$$I(\nu)=\frac{0.741l}{(-k)^{3/2}}+\frac{0.254l}{(-k)^{7/3}}-\frac{0.129l^3}{(-k)^{5/2}}+\frac{0.532l^2}{(-k)^{17/6}}-\cdots \qquad \text{(IV.18a)}$$

$$I(\nu)=\frac{0.507l}{k^{7/3}}-\frac{0.614l^2}{k^{17/6}}+\frac{0.320l^3}{k^{10/3}}-\cdots. \qquad \text{(IV.18b)}$$

If the first term in Eq. (IV.18a) is taken as the expression for the intensity in the red wing, agreement with the expressions of Kuhn[42] and Margenau is, of course, obtained. The earlier verification of Kuhn's expression for the intensity distribution in the red wing in the case of Na broadened by A as obtained by Kuhn[43] provide experimental justification for the Lindholm result.

Let us assume that the first term in Eq. (IV.18b) is sufficient to describe the intensity distribution in the violet wing. When a comparison of this expression with the measurements of Minkowski[70] for the violet wing of the Na D_2 line broadened by A is made, really excellent agreement is obtained. It might be mentioned here that, although the $(\Delta\nu)^{-\frac{3}{2}}$ dependence had been previously obtained for the red wing, the $(\Delta\nu)^{-7/3}$ dependence had certainly not been previously obtained for the violet wing. Minkowski had attempted to use a $(\Delta\nu)^{-2}$ dependence without the success which attended Lindholm's utilization of $(\Delta\nu)^{-7/3}$.

Kleman and Lindholm[41] experimentally investigated the broadening of Na lines by A and obtained excellent agreement in shift, half-width, and line contour with this theory.

The interruption theories to which we have so far devoted our attention may reasonably be classified as classical ones based on the Michelson application of the Fourier analysis to the problem. We shall show rather shortly that a quantum consideration also may lead to the mathematics of the Fourier analysis. Before doing so however, let us consider the interesting and important matters of distributions and detailed balancing which were studied by Van Vleck and Weisskopf and by Van Vleck and Margenau, respectively.

D. A Maxwell Distribution of Dipole Moments

Van Vleck and Weisskopf[88] were of the opinion that a revision of Lorentz' development, such that an agreement between this work and that of Debye be obtained was required. In essence this revision consisted in substituting for the random Lorentz distribution of x and $\dot{x}$ after a collision the Maxwell-Boltzmann distribution, and let us detail this.

The Lorentz solution to the equation of motion for the vibrating photoelectron is

$$x=\Re\left\{\frac{\epsilon a e^{i\omega t}}{m(\omega_0{}^2-\omega^2)}\left[1-\frac{1+(\omega/\omega_0)}{2\tau[-i(\omega_0-\omega)+1/\tau]}-\frac{1-(\omega/\omega_0)}{2\tau[i(\omega+\omega_0)+1/\tau]}\right]\right\} \quad \text{(IV.19a)}$$

$$=\Re\{ae^{i\omega t}(x'-ix'')\}. \quad \text{(IV.19b)}$$

In order to write down the desired result from Eq. (IV.19b), it is necessary to digress for a moment.

We begin by considering a molecule of electric dipole moment $\boldsymbol{\mu}$ which makes an angle ϑ with an external electric field $a\cos\omega t$. In this case now, we assume that after collision a Maxwell-Boltzmann distribution governs the dipole orientation (and hence, the orientation of x). If the last collision occurred at time t_0, the energy due to the dipole-field interaction is $\boldsymbol{\mu}\cdot\mathbf{a}$. Then we may obtain the mean polarization

$$\langle\mathfrak{M}\rangle=\frac{\mu\int_0^\pi \cos\vartheta\exp[\mu a\cos\vartheta\cos(\omega t_0)/kT]\sin\vartheta d\vartheta}{\int_0^\pi \exp[\mu a\cos(\omega t_0)/kT]\sin\vartheta d\vartheta}$$

$$\doteq\frac{\mu^2 a\cos(\omega t_0)\int_0^\pi \cos^2\vartheta\sin\vartheta d\vartheta}{kT_0\int_0^\pi \sin\vartheta d\vartheta}$$

$$=\frac{\mu^2 a}{3kT}\Re\{e^{i\omega_0 t}\} \quad \text{(IV.20)}$$

when we have expanded the exponential in a MacLaurin series and only retained the term containing a. The polarization per cubic centimeter may be obtained by averaging Eq. (IV.20) over the Lorentz distribution $a\exp(-t/\tau)$. The result is

$$P=\frac{Na\mu^2}{3kT}\Re\left\{\frac{e^{i\omega t}}{1+i\omega\tau}\right\}=a\Re\{(b_1-ib_2)e^{i\omega t}\}. \quad \text{(IV.21)}$$

The work done on the molecule by the radiation field will surely be given by the average value of $a\cos\omega t(dp/dt)$. Finally, if we divide this work by the energy flow in the field, $ca^2/8\pi$, we should determine the energy taken out of the field by the molecule, or, simply, the absorption coefficient.

Now

$$a\cos(\omega t)\frac{dp}{dt}=F\cos\omega t\left[\frac{Na\mu^2}{3kT}\frac{\omega^2\tau\cos(\omega t)}{1+\omega^2\tau^2}\right]$$

$$=\frac{Na\mu^2}{3kT}\frac{\omega^2\tau\cos(\omega t)}{1+\omega^2\tau^2} \quad \text{(IV.22a)}$$

and the average value is

$$\left\langle F\cos(\omega t)\frac{dp}{dt}\right\rangle=\frac{Na^2\mu^2}{6kT}\frac{\omega^2\tau}{1+\omega^2\tau^2}. \quad \text{(IV.22b)}$$

As a consequence we obtain the Debye result

$$\mathfrak{H}=\frac{\omega}{c}\frac{4\pi N\mu^2}{3kT}\frac{\omega\tau}{1+\omega^2\tau^2} \quad \text{(IV.23)}$$

for the absorption coefficient.

Now we note that the same result could have been obtained if we had set

$$\mathfrak{H}=\frac{4\pi b_2\omega}{c} \tag{IV.24}$$

from Eq. (IV.21). This tells us then that, from Eq. (IV.19b),

$$\mathfrak{H}=\frac{4\pi N\omega\epsilon x''}{c}=\frac{2\pi N\epsilon^2}{mc}\left(\frac{\omega}{\omega_0}\right)\times\left[\frac{1/\tau}{(\omega-\omega_0)^2+(1/\tau)^2}-\frac{1/\tau}{(\omega+\omega_0)^2+(1/\tau)^2}\right]. \tag{IV.25}$$

From the Lorentz theory we obtained Eq. (IV.25), the quantum analog of which is obtained by replacing $1/m$ by $(4/3)\pi/h\nu_{ij}|\mathbf{x}_{ij}|^2$. Next we introduce a Maxwell-Boltzmann distribution over the energy states and sum over these states with the result

$$\mathfrak{H}=\left(\frac{4\pi\nu N\epsilon^2}{3\hbar c}\right)\frac{\sum_j\sum_i|\mathbf{x}_{ij}|^2 f(\nu_{ij},\nu)e^{-E_j/kT}}{\sum_j e^{-E_j/kT}} \tag{IV.26a}$$

where the "shape factor" is given by

$$f(\nu_{ij},\nu)=\frac{1}{\pi}\left[\frac{(\delta/2)}{(\nu_{ij}-\nu)^2+(\delta/2)^2}-\frac{(\delta/2)}{(\nu_{ij}+\nu)^2+(\delta/2)^2}\right] \tag{IV.26b}$$

$$\delta=\frac{1}{\pi\tau}. \tag{IV.26c}$$

Let us now write

$$\sum_j\sum_i|\mathbf{x}_{ij}|^2 f(\nu_{ij},\nu)e^{-E_j/kT}$$
$$=\sum\sum_{j\neq i}[|\mathbf{x}_{ij}|^2 f(\nu_{ij},\nu)e^{-E_j/kT}+|\mathbf{x}_{ji}|^2 f(\nu_{ji},\nu)e^{-E_i/kT}]$$
$$=\sum_i\sum_j|\mathbf{x}_{ij}|^2 f(\nu_{ij},\nu)\frac{h\nu_{ij}}{2kT}e^{-E_j/kT}$$

since

$$\nu_{ji}=-\nu_{ij};\quad f(\nu_{ij},\nu)=-f(\nu_{ji},\nu);\quad |\mathbf{x}_{ij}|=|\mathbf{x}_{ji}|$$

and

$$(e^{-y}-e^{-x-y})\doteq\tfrac{1}{2}x(e^{-y}+e^{-x-y})\quad\text{for}\quad x\ll 1$$

so that Eq. (IV.19a) becomes:

$$\mathfrak{H}=\left(\frac{4\pi\nu N\epsilon^2}{6\hbar c}\right)\frac{h}{2kT}\frac{\sum_j\sum_i|\mathbf{x}_{ij}|^2\nu_{ij}f(\nu_{ij},\nu)e^{-E_j/kT}}{\sum_j e^{-E_j/kT}}. \tag{IV.27}$$

Now if Eqs. (IV.25) and (IV.27) were truly equivalent, as it would appear they should be, the latter would be expected to reduce to the former for $\nu_{ij}=0$. It is rather apparent that zero instead of such a reduction results. Van Vleck and Weisskopf remedied this discrepancy by using a Maxwell-Boltzmann distribution for x and $\dot{x}$ in the equations of motion for the vibrating photoelectron. Lorentz had used a random distribution of x and $\dot{x}$ in these equations.

A solution under these conditions yields, in place of Eq. (IV.19),

$$\langle x\rangle=\mathfrak{R}\left[\frac{\epsilon Fe^{i\omega t}}{m(\omega_0^2-\omega^2)}\left\{1-\frac{(\omega_0+\omega)(\omega/\omega_0^2\tau)}{2[1/\tau-i(\omega_0-\omega)]}+\frac{(\omega_0-\omega)(\omega/\omega_0^2\tau)}{2[1/\tau+i(\omega_0+\omega)]}\right\}\right]. \tag{IV.28}$$

In analogy to the method of obtaining Eq. (IV.25) we may then obtain

$$\mathfrak{H}=\frac{2\pi N\epsilon^2}{mc}\left(\frac{\omega}{\omega_0}\right)^2\times\left[\frac{1/\tau}{(\omega-\omega_0)^2+(1/\tau)^2}+\frac{1/\tau}{(\omega+\omega_0)^2+(1/\tau)^2}\right] \tag{IV.29}$$

of quantum equivalent

$$\mathfrak{H}=\left(\frac{4\pi\nu N\epsilon^2}{6\hbar c}\right)\frac{h}{kT}\frac{\sum_j\sum_i|\mathbf{x}_{ij}|^2\nu_{ij}f(\nu_{ij},\nu)e^{-E_j/kT}}{\sum_j e^{-E_j/kT}} \tag{IV.30a}$$

where

$$f(\nu_{ij},\nu)=\frac{1}{\pi}\frac{\nu}{\nu_{ij}}\left[\frac{\Delta\nu}{(\nu_{ij}-\nu)^2+(\Delta\nu)^2}+\frac{\Delta\nu}{(\nu_{ij}+\nu)^2+(\Delta\nu)^2}\right]. \tag{IV.30b}$$

Utilizing Eqs. (IV.30), we do indeed obtain Eq. (IV.23) for the case of zero resonant frequency, $\nu_{ij}=0$, and $\epsilon^2\sum|\mathbf{x}_{ij}|^2=\mu^2$ which at least is of import in resolving the paradox in the low-frequency region. Equation (IV.29) on the other hand, offers little change in the visible region, the $(\omega/\omega_0)^2$ factor yielding slightly more line asymmetry.

E. Detailed Balancing

The studies of Van Vleck and Margenau[87] did not serve to actually change the shape of the proposed absorption line, but they did finally show that the shapes of the spectral lines are the same in emission as in absorption, a point of no mean import. At any rate, since we accept the thesis that the integrated absorption

and emission balance each other in a Rayleigh-Jeans radiation field, detailed balance of emission and absorption results from this equivalence of absorption and emission line shapes. That is, a given frequency interval absorbs as much as it emits. The balance was obtained by a hitherto untried technique, that is, the summing of the work done on the oscillator by the electric field at collision as well as during the intercollision intervals. This yielded the same frequency by frequency power as that emitted between two collisions.

If our oscillator motion is described by $x(t) = x_0 \cos(\omega_0 t + \varphi)$ we may quite readily write down the Fourier analysis of $x(t)$. We may then integrate the expression for $x(\omega)$ and average $|x(\omega)|^2$ over a random distribution of φ to obtain a result which, when averaged over a random distribution of intercollision times, yields§§

$$|x(\omega)|^2 = \frac{x_0^2}{4\pi}\left[\frac{1}{a^2+(\omega_0-\omega)^2} + \frac{1}{a^2+(\omega_0+\omega)^2}\right] \quad \text{(IV.31)}$$

where now $a = 1/\tau$.

A point charge which is oscillating in one dimension radiates power of amount

$$\frac{2}{3}\frac{\epsilon^2}{c^3}\ddot{x}^2.$$

from the equation for simple harmonic motion $\ddot{x} = -\omega^2 x$. In addition, normalization of the Fourier components requires

$$\int_{-\infty}^{+\infty} [x(t)]^2 dt = \int_{-\infty}^{+\infty} |x(\omega)|^2 d\omega = 2\int_0^{\infty} |x(\omega)|^2 d\omega$$

so that

$$\ddot{x}^2 = -\omega^4 |x(t)|^2 = -\omega^4 |x(\omega)|^2.$$

We thus obtain for the power emitted by the oscillator in the frequency interval between ω and $\omega + d\omega$

$$P_E'(\omega)d\omega = \frac{2\epsilon^2}{3c^3} 2\omega^4 |x(\omega)|^2 d\omega. \quad \text{(IV.32a)}$$

This must be modified to include the intervals between all collisions over a long period t, however, since the Fourier analysis has only been carried out over one such period. As a result, Eq. (IV.32a) becomes

$$P_E(\omega)d\omega = \frac{atP_E'(\omega)d\omega}{t} = \frac{2}{3}\frac{\epsilon^2}{c^3} 2\omega^4 |x(\omega)|^2 a d\omega. \quad \text{(IV.32b)}$$

§§ In evaluating the Fourier integral we have taken the limits 0 and ϑ where ϑ is the time between two collisions, thus integrating $x_0 \cos(\omega t + \varphi)$ over the time between two collisions to obtain $x(\omega)$.

Substitution of Eq. (IV.31) into Eq. (IV.32b) then yields

$$P_E(\omega) = \frac{\epsilon^2 x_0^2}{3\pi c^3}\omega^4 \left[\frac{a}{a^2+(\omega_0-\omega)^2} + \frac{a}{a^2+(\omega_0+\omega)^2}\right]. \quad \text{(IV.33)}$$

Now in order to obtain detailed balancing, we must needs show that $P_A(\omega)$, the power absorbed at the frequency ω, corresponds to this.

If the velocity proportional viscous drag force $g\dot{x}$ is dropped from the electronic equation of motion, the Van Vleck-Margenau equation is obtained as

$$\ddot{x} + \omega_0^2 x = \frac{\epsilon F}{m}\cos(\omega t + \varphi)$$

a solution to which is sought under the Lorentz boundary conditions, $x = \dot{x} = 0$.

The solution in question leads to the following result for the work done—per unit time—by the field on the electron *between* collisions:

$$\frac{W}{t} = \frac{\epsilon^2 F^2}{4m}\left[\frac{a}{a^2+(\omega-\omega_0)^2} + \frac{a}{a^2+(\omega+\omega_0)^2}\right]. \quad \text{(IV.34)}$$

The work done—per unit of time—by the field *at* collision may be obtained as

$$\frac{\epsilon^2 F^2 a}{4m\omega}\left[\frac{\omega-\omega_0}{a^2+(\omega-\omega_0)^2} - \frac{\omega+\omega_0}{a^2+(\omega+\omega_0)^2}\right] \quad \text{(IV.35)}$$

which, when added to Eq. (IV.34), yields

$$P_A(\omega) = \frac{W}{t} = \frac{\epsilon^2 F^2 \omega}{4m\omega_0} \times \left[\frac{a}{a^2+(\omega-\omega_0)^2} - \frac{a}{a^2+(\omega+\omega_0)^2}\right]. \quad \text{(IV.36)}$$

A consideration of Eqs. (IV.33) and (IV.36) is sufficient to show that detailed balance has not been obtained if for no other reason than that the shape factors as given by the brackets in the two equations differ. Van Vleck and Margenau overcame this difficulty by passing from the Lorentz to the Van Vleck-Weisskopf boundary conditions on x and $\dot{x}$ at collision. This has the effect of adding the term

$$\frac{\epsilon^2 F a}{4m\omega^2}\left[\frac{\omega^2-\omega\omega_0}{a^2+(\omega-\omega_0)^2} + \frac{\omega^2+\omega\omega_0}{a^2+(\omega+\omega_0)^2}\right] \quad \text{(IV.37)}$$

to Eq. (IV.37) to arrive at the power absorbed from the light wave as a function of frequency.

If we let $\Upsilon(\omega)d\omega$ be the energy density in the field for the frequency interval ω to $\omega + d\omega$, we may replace F^2 in

Eqs. (IV.36) and (IV.37) by

$$F^2 = \frac{4}{3}\frac{m\omega^2\omega_0^2}{\pi c^3}x_0^2 \quad \text{(IV.38)}$$

in the sum of Eqs. (IV.36) and (IV.37) to obtain

$$P_A(\omega) = \frac{\epsilon^2 x_0^2}{3\pi c^3}\omega^4\left[\frac{a}{a^2+(\omega_0-\omega)^2}+\frac{a}{a^2+(\omega_0+\omega)^2}\right]. \quad \text{(IV.39)}$$

A comparison of Eqs. (IV.33) and (IV.39) shows that a detailed balance condition between absorption and emission has indeed been obtained by considering, in absorption, the work done by the field (a) *between* collisions and (b) *at* collisions and by including the Maxwell-Boltzmann boundary conditions.

We have remarked that the boundary conditions especially are an approximation. In an actual atomic system, of course, finite collision times would have to be considered. In addition, the Planck radiation law would more logically replace the Rayleigh-Jeans law. In the limit of low frequency, however, the former reduces to the latter, so that, as in the case of the earlier Van Vleck-Weisskopf considerations, the results are particularly applicable to the microwave region.

F. The Correlation Function Treatment

At the beginning of the present chapter we remarked on the various methods which have been utilized to demonstrate the quantum justification of the use of the Fourier analysis in line broadening. We shall now give an extremely simple demonstration of such a justification by using the state growth equations for transitions accompanied by the emission of radiation.

Let us suppose that the natural width of the spectral line is small compared to the interruption width so that we may neglect the radiation damping constant, γ in Eq. (II.24). Then this equation may be approximately written down as

$$b_i(t) = -i\,\text{const}\int_0^t dt'$$

$$\times\exp\left\{\frac{i}{\hbar}\int_0^{t'}[E_B(t'')-E_A(t'')]dt''+\omega_i t\right\}. \quad \text{(IV.40)}$$

We have previously considered the argument which results in the relation between state growth coefficient and line intensity distribution:

$$I(\omega) = |b_\omega(\infty)|^2 = \int_0^\infty dt'\int_0^\infty dt''$$

$$\times\exp\left[-i\int_{t''}^{t'}P(t)dt+i\omega(t'-t'')\right] \quad \text{(IV.41)}$$

where now $P(t)$ is the angular frequency separation of the perturbed energy levels.

We will agree that Eq. (IV.41) is a form of the Fourier integral which we have previously encountered, for example, Eq. (IV.4). Having obtained this Fourier integral after a quantum fashion, we should now feel justified, as did Foley,[17] in proceeding with the treatment of what this latter author was pleased to dub the "correlation function."

In connection with the evaluation of Eq. (IV.4) we mentioned the replacement of the second integral by the statistical average. The principal behind such a procedure had, as we have remarked, been set forth by Lenz. The exponential factor [which we now designate as $\varphi(t_0)$] was named by Foley. The remainder of the work which has been done on the interruption theory has consisted of the attempts of Foley[17] and Anderson[2] to evaluate this correlation function, a term which appears to have been given a favorable reception. Let us sketch Foley's evaluation and turn to that of Anderson which does seem a bit more general of application.

We write:

$$\varphi(t_0) = \langle y(t_0)\rangle = \int_0^\infty dt''\exp\left[i\int_{t''}^{t''+t_0}P(t)dt\right]$$

$$= \left\langle\exp\left[i\int_0^{t_0}P(t)dt\right]\right\rangle. \quad \text{(IV.42)}$$

Much after the Lindholm fashion we may replace the integral within the exponent by a sum over phase shifts:

$$\sum_{i=1}^{n}\eta_i = \int P(t)dt$$

which we now average over the distribution of phase shifts, $p(\eta)$:

$$\int dt''\exp[-i\sum_{i=1}^{n}\eta_i]$$

$$= \prod_{i=1}^{n}\int_{-\infty}^{+\infty}d\eta\,p(\eta)(\cos\eta-i\sin\eta) = (A-iB)^n. \quad \text{(IV.43)}$$

Finally this result is averaged over the standard distribution of intercollision times,

$$\varphi(t_0) = \sum_{n=0}^{\infty}\frac{e^{-\vartheta/\tau}}{n!}\left(\frac{\vartheta}{\tau}\right)^n(A-iB)^n$$

$$= \exp\left[\frac{A-iB}{\tau}\vartheta-\frac{\vartheta}{\tau}\right] \quad \text{(IV.44)}$$

which, when substituted into Eq. (IV.41), results in

$$I(\omega) = \frac{(1-A)/\pi\tau}{[(1-A)/\tau]^2+[B/\tau-\omega]^2}. \quad \text{(IV.45)}$$

The line shift and half-width for this intensity distribution are then

$$\Delta = B/\tau \qquad \text{(IV.46a)}$$

$$\delta = \frac{2(1-A)}{\tau}. \qquad \text{(IV.46b)}$$

Equation (IV.43) provides the means for the requisite evaluation of the coefficients A and B, a somewhat lengthy procedure which we shall not detail. With which we turn our attention to Anderson's work on the subject.

Margenau[58] had early (1935) demonstrated that the Fourier integral will reduce to the Margenau statistical distribution for sufficiently low velocities. Now Anderson, armed with the Lenz method of evaluation, has carried this procedure a step further. He has (1) obtained the statistical distribution for high pressures and for the line wing, (2) obtained the simple interruption result for low pressures and near line center, and (3) numerically evaluated the intermediate cases.

We begin by supposing that the perturbations arising due to the various members of our gaseous assembly are additive so that

$$\Delta\omega = \sum_{i=1}^{n} \Delta\omega(R_i). \qquad \text{(IV.47)}$$

This means that Eq. (IV.42) assumes the form

$$\varphi(t_0) = \left\langle \exp\left[i\int_0^{t_0} \nu(t)dt \right] \right\rangle = \left\langle \prod_{i=1}^{n} \exp\left[i\int_0^{t_0} \nu(R_i)dt \right] \right\rangle$$

$$= \left\langle \exp\left[i\int_0^{t_0} \nu(R)dt \right] \right\rangle^n. \qquad \text{(IV.48)}$$

Now let us remark in connection with this equation that the replacement of the average of the product by the product of n identical averages is predicated on (1) the identical nature of all broadeners and (2) the absence (assumed) of interactions between these broadeners. No justification appears required for

$$R^2 = (x_0 + \langle v\rangle t)^2 + \rho^2.$$

This means, however, that the averaging process will now be one over x_0 and ρ

$$\varphi(t_0) = \frac{1}{V^n}\left\{ \int_0^\infty 2\pi\rho d\rho \int_{-\infty}^{+\infty} dx_0 \exp\left[i\int_0^{t_0} \nu(R)dt \right] \right\}^n$$

$$= \left[1 - \frac{2\pi N}{n}\int_0^\infty \rho d\rho \int_{-\infty}^{+\infty} dx_0 \times\left(1 - \exp\left[i\int \nu(R)dt \right] \right) \right]^n$$

$$= \exp[-NV'(t_0)] \qquad \text{(IV.49a)}$$

since

$$\lim_{n\to\infty}\left[1 - \frac{A}{n} \right]^n = \exp[-A]$$

and where

$$V'(t_0) = 2\pi\int_0^\infty \rho d\rho \int_{-\infty}^{+\infty} dx_0 \times\left\{ 1 - \exp\left[i\int_0^{t_0} \nu([(x_0+\langle v\rangle t)^2+\rho^2]^{\frac{1}{2}})dt \right] \right\} \qquad \text{(IV.49b)}$$

with $V = n/N$.

Now in order to obtain the low velocity statistical limit, one simply lets $\langle v\rangle = 0$ with the desired result. A comparison with Eq. (III.8) serves to demonstrate this quite nicely.

For the simple interruption theory limit we suppose the reverse situation to prevail, namely, $\langle v\rangle$ large compared to x_0, thus justifying ignoring this latter coordinate in the exponential. It would seem intuitively apparent that the range of integration $-\langle v\rangle\tau$ to $+\langle v\rangle\tau$ would make the greatest contribution to the integral in x_0. Such a limitation on the range of integration of x_0 leads to

$$\varphi = \langle v\rangle\tau \int_0^\infty \rho d\rho \left(1 - \exp\left[i\int_{-\infty}^{+\infty} \nu([\rho^2 + \langle v\rangle^2 t^2]^{\frac{1}{2}})dt \right] \right)$$

$$= (\sigma_r + \sigma_i)\langle v\rangle\tau = \sigma\langle v\rangle\tau \qquad \text{(IV.50a)}$$

from which one obtains [through Eq. (IV.58)] for the intensity distribution

$$I(\omega) = \frac{N\langle v\rangle\sigma_r}{(\omega-\omega_0-N\langle v\rangle\sigma_i)^2 + (N\langle v\rangle\sigma_r)^2}. \qquad \text{(IV.50b)}$$

This should suffice to demonstrate the fashion in which the theory reduces in the two limiting cases. For the intermediate values numerical calculation is required. It is understood that Anderson has carried out such calculations, but the results have not as yet been published.

V. RESONANCE BROADENING

A. The Qualitative Basis of Self-Broadening

In our specific studies of the last three chapters the broadening phenomena considered have resulted from the presence of other atoms. Although different interatomic forces have been utilized as agents for this broadening, no direct consideration has been given to the relation between the emitter and the broadener, that is, no consideration has been given to whether the broadeners are of the same or of a different species than the emitter. In this chapter we propose to devote ourselves to that unique set of phenomena which occur when the broadeners are the same type atoms as the

emitter, as opposed to the various broadenings which may occur when the broadeners are of a different type from the emitter.

In 1915 Rayleigh[74] appears to have first raised the question as to the possible difference in broadening effects in the presence of like atoms as broadeners and in the presence of foreign gas atoms. When foreign and self-broadening were subsequently studied as distinct phenomena, it was early recognized that the magnitude of the latter was much greater than that of the former. We may rather easily present a qualitative picture of these self broadening (a term which we shall subsequently replace by resonance broadening) effects either from the classical or quantum viewpoint. Since such is the case, let us begin with the classical explanation of the phenomenon.

In the classical sense, the similar atoms of a radiating gas are made up of a collection of like oscillators of the same natural frequency. Because of the sameness of this natural frequency, a strong coupling of these oscillators may be expected to occur in the usual classical manner given some basic, say electrostatic, coupling force. This in turn will cause a spread of the coupled oscillator frequencies about the natural frequency of a single oscillator.

From a quantum-mechanical sense, the same strong broadening can be inferred. Let us hypothesize a two-atom system in which the level degeneracy arising from the sameness of the atoms results in an energy perturbation dependent on the inverse cube of the atomic separation. We recall that under the same interaction force two unlike atoms give rise to an inverse sixth power dependence on the atomic separation. Thus, obviously greater broadening will result from the presence of like molecules. We shall later discuss the reduction in state lifetime for the case of resonance broadening which results in a broadening of the spectral line, but the state degeneracy which we have mentioned above we shall allow to suffice as an introductory consideration.

B. The Coupled Oscillator Theory

The first attack on the problem of resonance broadening was made by Holtsmark[31] in 1925. This author considered the problem from a classical point of view which depicted the atoms of the absorbing, self-broadening gas as classical oscillators.

Now we suppose a coupling force, dependent on the electric dipole moment of the atoms, to be present. Let us begin by considering two of these classical oscillators each consisting primarily of a "quasi-elastically bound" electron. If x_1 is the vibrational coordinate of the first oscillator and x_2 that of the second, and these two coordinates are coupled through the potential term $k_3x_1x_2$, then the frequencies of the coupled system are surely

$$\omega^2=\tfrac{1}{2}\left(\omega_1^2+\omega_2^2\pm\left[(\omega_1^2-\omega_2^2)+4\left(\frac{k_3}{m}\right)^2\right]^{\frac{1}{2}}\right). \quad \text{(V.1)}$$

If the two atoms are unlike then $|\omega_1^2-\omega_2^2|\gg k_3/m$ and we obtain

$$\omega^2-\omega_{1,2}^2=\pm\frac{k_3^2}{m^2(\omega_1^2-\omega_2^2)} \quad \text{(V.2)}$$

whereas, in the case $\omega_1=\omega_2=\omega_0$ (like atoms), we obtain

$$\omega^2-\omega_0^2=\pm\frac{k_3}{m}. \quad \text{(V.3)}$$

A consideration of Eqs. (V.2) and (V.3) serves to illustrate the much larger frequency shift accompanying the coupling of like oscillators since k_3 is always small. This essentially forms the basis for Holtsmark's considerations. Let us first determine k_3, and then proceed with the problem.

We have assumed our coupling force to arise from the electric dipole moments of the oscillators, and we may write the dipole potential as

$$U=\mathfrak{F}_i\cdot\boldsymbol{\mu}_k=\frac{\epsilon^2}{r^3}[1-3\cos^2\gamma_{ik}]x_ix_k, \quad \text{(V.4)}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_k$ are the dipole moments of the ith and kth atom.

γ_{ik} is now the angle between either $\boldsymbol{\mu}_i$ or $\boldsymbol{\mu}_k$ and $\mathbf{r}$. We further have supposed $\boldsymbol{\mu}_i \| \boldsymbol{\mu}_k$.

Then the force on atom i due to the field produced by atom k is, since $\mathfrak{F}=-\nabla U$

$$\mathfrak{F}=\frac{\epsilon^2u_{ik}x_k}{r_{ik}^3}=ma_{ik}x_k \quad \text{(V.5a)}$$

with

$$u_{ik}=(1-3\cos^2\gamma_{ik}). \quad \text{(V.5b)}$$

Now if we consider a system of n atoms instead of our original two-atom system we may obtain

$$\ddot{x}_i+\omega_0^2x_i+\sum_k a_{ik}x_k=0 \qquad \begin{matrix} k\lesseqgtr i \\ i=1,2,\cdots,n. \end{matrix} \quad \text{(V.6)}$$

From Eq. (V.6) we evolve the secular determinant for the problem

$$\begin{vmatrix} \Lambda & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & \Lambda & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & \Lambda & \cdots & a_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \Lambda \end{vmatrix}=0 \quad \text{(V.7)}$$

where the abbreviation $\Lambda=\omega^2-\omega_0^2$ has been utilized.

Now it is rather obvious that it would be somewhat difficult at best to obtain the roots of the determinant of Eq. (V.7) for the general case, in fact, it presently appears impossible. We can, however, determine the mean and the root-mean-square spread of the roots of this equation after a fashion which we shall not detail. After Holtsmark had done precisely this, he proceeded

to make an assumption which does not appear too well justified in the light of subsequent—and one might add previous—knowledge of resonance broadening. Quite simply he supposed the spectral line to be of a Gaussian shape. In the extreme wings of the line this appears to be a reasonable supposition, and we might recall that earlier Holtsmark had been required to assume such a shape in the wings of the electric field strength distribution in connection with the Stark broadening. The connection between the two situations is certainly obscure, however, and subsequent studies of resonance broadening seem to confirm the fact that the Gaussian distribution is not the proper one. At any rate, such an assumption leads to

$$I(\omega')=\frac{1}{[2\pi\langle\omega'^2\rangle]^{\frac{1}{2}}}\exp(-\omega'^2/\langle\omega'^2\rangle) \qquad (V.8a)$$

of half-width

$$\delta=2\omega'=2.36[\langle\omega'^2\rangle]^{\frac{1}{2}}=103.5\times10^7\frac{\sqrt{N}}{\rho^{\frac{3}{2}}}\frac{1}{2\omega_0} \qquad (V.8b)$$

in which $\langle\ \rangle$ indicates the mean value of the quantity enclosed.

The feature in the half-width equation which appears most striking—although perhaps not most pleasingly so—is the dependence on the root of the density. Other theories as well as available experimental evidence appear to be unanimous in their refutation of any such dependence claim. A consideration of the half-width equation will probably indicate to the reader, however, that ρ offers a bonnet out of which we might possibly produce a different pressure dependence. Schutz-Mensing[66] has produced just such a dependence in what seems a very straightforward manner.

We take the separation of the emitter and the ith broadener as ρ_i, and proceed to average over the ρ_i. Schutz-Mensing obtained a value $0.55N^{-\frac{1}{3}}$ for $\langle\rho\rangle$. It is obvious that such a value for ρ in Eq. (V.8b) results in a linear density dependence for the half-width. For the root mean square spread of the energies—resulting from the roots of the determinant—this author obtained

$$[\langle\omega'^2\rangle]^{\frac{1}{2}}=\left[\frac{24\pi}{15}\right]^{\frac{1}{2}}\frac{1}{\omega_0}\left(\frac{\epsilon^2}{m}\right)N. \qquad (V.9)$$

C. The Statistical Resonance for Binary Encounters

Let us assume that one member of our two like particle system will proceed from a state of energy E_1 to the ground state of energy E_0 with accompanying radiation. Then familiar degenerate perturbation theory yields, as the energy of this interaction

$$E_1^{(1)}=h(\Delta\nu)=\begin{cases}-\dfrac{2}{3}\dfrac{\epsilon^2}{r^3}|r_{12}|^2 & \text{for } m=0\\[2ex] +\dfrac{1}{3}\dfrac{\epsilon^2}{r^3}|r_{12}|^2 & \text{for } m=\pm1\end{cases} \qquad (V.10a)$$

where, as usual, m is the magnetic quantum number, and where

$$|r_{12}|^2=\frac{3h}{8\pi^2m\nu_0}f_{12}; \qquad (V.10b)$$

all of which leads to

$$\Delta\nu=\begin{cases}-\dfrac{\epsilon^2}{r^3}\dfrac{2}{8\pi^2m\nu_0}f_{12} & \text{for } m=0\\[2ex] +\dfrac{\epsilon^2}{r^3}\dfrac{1}{8\pi^2m\nu_0}f_{12} & \text{for } m=\pm1.\end{cases} \qquad (V.11)$$

From Eqs. (V.3) and (V.4) the classical result,

$$\Delta\nu=\pm\frac{\epsilon^2}{r^3}\frac{(1-3\cos^2\gamma_{ik})}{4\pi^2m(\nu+\nu_0)} \doteq\pm\frac{\epsilon^2}{r^3}\frac{(1-3\cos^2\gamma_{ik})}{8\pi^2m\nu_0}, \qquad (V.12)$$

may be written down for comparison.

Equation (V.11) may be rewritten as

$$E_1^{(1)}=h\nu'=\gamma\left(\frac{\epsilon^2hf_{12}}{8\pi^2m\nu_0}\right)\frac{1}{r^3} \qquad (V.13)$$

where γ is the statistical weight factor associated with m. After the fashion of Margenau and Watson, let us equate γ to unity. We consider this an averaging process over the possible atomic orientations. The result is

$$U=E_1^{(1)}=ar^{-m}=\left(\frac{\epsilon^2hf_{12}}{8\pi^2m\nu_0}\right)r^{-3} \qquad (V.14a)$$

so that

$$a=\frac{\epsilon^2hf_{12}}{8\pi^2m\nu_0}. \qquad (V.14b)$$

From Eq. (III.15) the half-width of the resonance broadened spectral line is

$$\delta=\frac{2U_0}{h}=\frac{4}{3}\pi^2Na,$$

so that

$$\delta=\frac{\epsilon^2f_{12}}{6m\nu_0}N. \qquad (V.15)$$

Equation (V.15) was originally obtained by Margenau and Watson[95] in a slightly different manner.

D. The Simple Interruption Result

As we recall the simple interruption theory prognosticates the same spectral line form regardless of the interaction forces which give rise to the broadening under consideration. Thus, Eqs. (IV.1) and (IV.2) for

the line shape and half-width remain unchanged by virtue of the resonance forces. Now this is, of course, not the case for the optical collision diameter ρ for it will be given by Eq. (IV.3). Let us specificize.

The resonance interaction may be expressed in the form

$$\Delta(r)=K/r^3$$

which means that the phase change is

$$\eta=\int_{-\infty}^{+\infty}\frac{Kdt}{[(\langle v\rangle^2t^2+\rho^2)^3]^{\frac{1}{2}}}=\frac{K}{\langle v\rangle\rho^2}\int_{-\infty}^{+\infty}\frac{dx}{[(x^2+1)^3]^{\frac{1}{2}}}$$

$$=\frac{2K}{\langle v\rangle\rho^2}=1. \qquad (V.16)$$

From Eq. (V.16) we obtain

$$\rho=\left[\frac{2K}{\langle v\rangle}\right]^{\frac{1}{2}} \qquad (V.17)$$

which is here expressed in angular frequency units ω. Now Eq. (V.13) tells us that

$$K=\epsilon^2/2m\omega_0. \qquad (V.18)$$

The substitution of Eqs. (V.17) and (V.18) yields the optical collision diameter from which we may obtain the half-width as

$$\delta=2\pi\frac{\epsilon^2}{m\omega_0}f_{nm}N. \qquad (V.19)$$

E. Qualitative Consideration of the Energy Transfer Theory

Some four years after the publication of Weisskopf's theory Furssow and Wlassow[19] entered the field with what did amount to a new conceptual approach. Their entry was marked by the intriguing statement that the Weisskopf theory—and the Lenz theory in passing—was completely incorrect.‖‖ It would appear reasonable for us to briefly consider the reasons which they advance for this conclusion.

For greatest accuracy of consideration let us quote the Furssow and Wlassow statement verbatim before attempting to consider their reasoning.

"To be sure Weisskopf's considerations of the collisions between the like atoms is not entirely correct. Weisskopf without foundation applies the correct concept of the mechanism of the collision damping in the case of nonextinguishing atoms of different types to the case of like atoms. As is known the broadening of the lines through collisions result not only when the wave train emitted by the atom is propagated after the collision (extinguishing gases) but also at the time when at the collision a change of the vibration phase of the excited atoms sets in (nonextinguishing gases). In order to compute the phase shift during the collision one must take the *change of frequency* $\Delta\omega$¶¶ of the emitting atom which occurs through the interaction with foreign atoms *and integrate over the collision time.*¶¶ If one makes this integral equal to one then the magnitude of the optical collision diameter can be evaluated. The collisions of the like atoms Weisskopf takes instead of arbitrarily the difference between the frequency of one of the normal vibrations of a system of two dipole linked like linear oscillators and the frequency of the isolated atom. It is clear that this computation is based on a misunderstanding. $\Delta\omega$ is in terms of its nature the change of the frequency of the emitting atom, the light of which is analyzed according to Fourier. That however which Weisskopf substitutes instead of it has nothing at all to do with the matter because of the degeneracy . . . the concept that in the collision of two like atoms one of which is excited and the other of which is unexcited the vibration phase of the excited atom is changed does not at all correspond to reality."[19]

It might be remarked first that the $\Delta\omega$ which they believe to have been improperly chosen seems to be the very one whose choice they advocate for the following reason ". . . *take the change of the frequency* $\Delta\omega$ . . . and *integrate over the collision time.*"[19] Now it is true that integration is not extended over these limits in Eq. (V.16), but let us recall that these limits are a good approximation which is predicated on the reasonable assumption that no appreciable contribution to the integral arises outside the proper limits. If they base their argument on the "degeneracy"—they do not define this degeneracy but we can assume that it is that degeneracy arising from the indistinguishability of the atoms of our ensemble—then the question of which approach is more nearly correct has to be answered a little more conclusively than has been done to date. Surely though the simple interruption theory has not been so devastated by the above quoted argument that it cannot be considered as remaining a reasonable approximation on proper application.

The Furssow and Wlassow theory really has its basis in a conception advanced by Kallmann and London[41] in 1929 to the effect that like atoms may simply exchange energy between each other without accompanying radiation. Even a cursory consideration of this conception gives an inkling of an application to the theory of self-broadening. We may examine it from either a quantum or a classical viewpoint.

First the classical viewpoint. To begin with we consider our atoms as classical oscillators, and again we have the dipole interaction between two like oscillators, one of which is excited and the other of which is unexcited. As a result of this interaction the amplitude of oscillation of the excited atom decreases while that of

‖‖ Houston took mild exception to this statement to the effect that ". . . the criticism of Weisskopf's work contained in this paper does not appear to be justified . . ."[32] but did not reply to the Furssow and Wlassow criticism specifically.

¶¶ Italics added.

the unexcited atom increases, that is, an energy transfer occurs. This is effectively the same as damping the oscillations of the excited atom, and, as in the case of the Lorentz damping will result in a broadening of the spectral line. Let us now consider qualitatively the quantum-mechanical explanation of this effect.

Energy is assumed to be transferred as a result again of the dipole interaction from an excited to an unexcited atom of the same kind. Whatever the lifetime of the excited state would be under conditions of no transfer this lifetime will be greatly shortened by a high probability of transfer before radiation. Now we may recall that one form of the Heisenberg uncertainty principle states that $\Delta E \Delta t > h$. Thus, if, as is the case in the ground state, the state lifetime is infinite, the ground state will be infinitely well defined or virtually infinitely narrow. As soon as we consider a state with a finite lifetime, however, the situation changes. A certain lifetime Δt will give us a certain indefiniteness ΔE in the state energy, or a certain level width. This level width in turn will mean that a spectral line arising from the combination of this level with another will be broadened as a consequence. If we decrease our state lifetime by this energy transfer, quantum mechanics decrees that we indirectly broaden our spectral line by widening the energy level.

These then are the classical and quantum forms of the theory as advanced by Furssow and Wlassow. It now remains only to rewrite our qualitative conjectures after a quantitative fashion. Let us do so rather briefly.

F. The Classical Energy Transfer (Low Pressure)

To begin with it is of course necessary to make a few simple assumptions regarding our system. In this consideration we are assuming our atom to be a classical harmonic oscillator. Now we shall consider that the electronic transition which gives rise to our broadened spectral line proceeds to the ground state. We will assume that only one valence electron is responsible for the spectral line under consideration. Finally let us suppose that the excited atom, of which our system contains one for our purposes here, moves in the neighborhood of the remainder of the unexcited atoms. The excited atom we then consider as moving rectilinearly with velocity v. We might possibly bring up the same type of objection to the utilization of a linear velocity here as was brought up by Jablonski in objection to Weisskopf's utilization of a linear velocity in a central force problem. However this is probably minor at this stage.***

In considering our interaction which leads to the self-broadening of the line, we shall assume this interaction to take place between our emitter and one of the unexcited broadening atoms. The interaction in question we presume to be due to the electric dipole. Lagrange's equations may then be solved by a method of successive approximations to obtain

$$x_1^{(1)} = \frac{3A}{2\omega_0} \int^t \Lambda \cos\alpha \cos\gamma dt \sin\omega_0 t$$

$$y_1^{(1)} = \frac{3A}{2\omega_0} \int^t \Lambda \cos\beta \cos\gamma dt \sin\omega_0 t \qquad \text{(V.20)}$$

$$z_1^{(1)} = -\frac{A}{2\omega_0} \int^t \Lambda (1 - 3\cos^2\gamma) dt \sin\omega_0 t$$

$$x_2^{(1)} = y_2^{(1)} = 0; \quad z_2^{(1)} = A \cos\omega_0 t.$$

Equations (V.20) are quite enlightening in that they serve to show the manner in which the energy is transferred from the initially excited atom to the initially unexcited one during the course of an optical collision.

Let us assume that the optical collision, which serves to transfer energy from the excited to the unexcited atom, lasts a time interval τ_0. If this is the case, and if we assume the collision to begin at time $t=0$, then the collision must be initiated at a distance $-v\tau_0/2$ from ρ.

Then after a collision of duration τ_0, the amplitude, for example, of $x_1^{(1)}$ we may obtain from Eq. (V.20)

$$X_1^{(1)} = \frac{3A}{2\omega_0} \int_0^{\tau_0} \Lambda \cos\alpha \cos\gamma dt. \qquad \text{(V.21)}$$

It is apparent that $\alpha\beta\gamma$ are time dependent through the heat motion of the atoms. We may express these quantities as functions of time and integrate to obtain

$$X_1^{(1)} = \frac{\epsilon^2}{m\omega_0} \frac{A}{\rho^2 v} (2\cos\alpha_2 \cos\gamma_2 + \cos\alpha_1 \cos\gamma_1), \text{ etc.} \qquad \text{(V.22)}$$

We are now desirous of obtaining the amount of energy which is transferred during the collision from the initially excited atom to the initially unexcited one. In the present case the energy E which has been transferred to the initially unexcited atom will be given by the sum of the squares of the amplitude components as given by Eq. (V.22) multiplied by $\frac{1}{2}m\omega_0^2$. We then obtain for the transferred energy

$$\mathcal{E} = \frac{\epsilon^4}{m^2\omega_0^2} \frac{1}{\rho^4 v^2} \sin^2\gamma_1 E \qquad \text{(V.23)}$$

where E, the initial energy of the excited molecule is, of course, $\frac{1}{2}\omega_0^2 A^2$.

Equation (V.23) is predicated on the first approximate solutions as given by Eq. (V.20). The validity of these Eqs. (V.20) is assumed only for large transit distances. Furssow and Wlassow defined the minimum distance of closest approach—transit distance—as that distance at which the transferred energy is equal to the energy at time zero of the initially excited oscillator,

*** *Most authors appear to consider it minor at any stage, although this is not conclusive.*

that is, $\mathcal{E}=E$. Thus if ρ_0 is taken as the minimum transit distance Eq. (V.23) leads us to the following criteria for large distances

$$1/\rho_0^2 v = m\omega_0/\epsilon^2. \qquad \text{(V.24)}$$

In order to determine the total energy which our initially excited atom will lose in the course of time, it will be necessary to ascertain the total loss of energy to all the atoms in the neighborhood. In passing through a layer of thickness Δs lying perpendicular to the direction of motion of the excited atom to those atoms a distance ρ_0 or more away the emitter will lose energy of amount

$$dE = -2\pi N\Delta s\int_{\rho_0}^{\infty}\mathcal{E}\rho d\rho = -\pi N\Delta s\frac{\epsilon^4}{m^2\omega_0^2}\sin^2\gamma_1\frac{1}{\rho_0^2 v^2}E$$

$$= -\pi N\frac{\epsilon^2}{m\omega_0}\sin^2\gamma_1 E dt \qquad \text{(V.25)}$$

from Eq. (V.24) and $\Delta s = vdt$. From Eq. (V.25)

$$E = E_0 e^{-\gamma t} \qquad \text{(V.26a)}$$

where

$$\gamma = \frac{2\pi}{3}\frac{N\epsilon^2}{m\omega_0}. \qquad \text{(V.26b)}$$

From this knowledge of the attenuation of the initially excited atom's energy we would like to determine the line broadening.

The electric vector of the radiation field will be of the form:

$$\mathbf{F} = \mathbf{F}_0 e^{-\gamma t/2} e^{i\nu_0 t}.$$

In the usual manner the electric vector may be expanded in a Fourier integral from which may be obtained

$$I(\nu) = |F(\nu)|^2 \propto \text{const}\frac{(\gamma/2)}{(\nu-\nu_0)^2 + (\gamma/2)^2} \qquad \text{(V.27)}$$

the half-width of the emitted spectral line is found to be

$$\delta_R' = \frac{2\pi}{3}\frac{\epsilon^2}{m\omega_0}N. \qquad \text{(V.28)}$$

It is then apparent that the effect of this energy transfer at large transit distances is a damping of the atomic oscillations resulting in a broadening of the spectral line whose half-width is given by Eq. (V.28).

For the close approaches they simply took the half-width as given by the simple interruption theory:

$$\delta_I = 2\pi\rho_0^2 N\langle v\rangle = 2\pi\frac{\epsilon^2}{m\omega_0}N. \qquad \text{(V.29)}$$

The total half-width is then

$$\delta = (\delta_R' + \delta_I')f = \frac{8\pi}{3}\frac{\epsilon^2}{m\omega_0}Nf. \qquad \text{(V.30)}$$

The manner in which the dipole interaction between two like atoms causes a transfer of the oscillatory energy from the initially excited atoms to the remaining initially unexcited atoms has been shown. This energy transfer then acts as a damping force on the oscillatory motion of the atom and, as a result, the emitted radiation is broadened into a spectral line of finite width. We shall see that a quantum consideration under certain specific assumptions leads to the same results.

G. Quantum Treatment of Low-Pressure Self-Broadening

We have sketched qualitatively the quantum-mechanical theory of resonance broadening by energy transfer, and our first task in a quantitative consideration will be to ascertain the time change of eigenfunctions—probability amplitudes—of the excited atom due to this energy transfer.

Again the system is initially taken as two like atoms one of which is excited and one of which is unexcited. The motion of the atom will be considered classically as is normally done in problems of this kind with quite reasonable justification. The potential of interaction between the two atoms is still the dipole potential. The symbols such as ρ, α_1, etc. which were utilized in the classical consideration will again appear with the same connotation.

Familiar procedures of degenerate perturbation theory may be used in obtaining the state growth coefficients.

$$a_1 = a; \quad a_2 = a_3 = 0$$

$$b_1 = \frac{a}{i\hbar}\int_0^{\tau_0} U_{n00;\, m10}^{m10;\, n00}dt, \text{ etc.} \qquad \text{(V.31)}$$

The a_i refer to the states of the initially excited while the b_i refer to those of the initially unexcited atom. In finding a matrix element of U we carry the angle functions as constants—it may be recalled that these angle functions depend on the heat motion of the atoms. Finally then Eq. (V.31) becomes

$$b_1 = \frac{a}{i\hbar}\epsilon^2(n00|z|m10)^2\int_0^{\tau_0}\frac{(1-3\cos^2\gamma)}{R^3}dt, \text{ etc.} \qquad \text{(V.32)}$$

Again, as in the classical case, we allow τ_0 (the collision time) to approach infinity and integrate Eq. (V.32).

It is apparent that the probability for excitation of the initially unexcited atom is

$$|b_1|^2 + |b_2|^2 + |b_3|^2 = a^2\frac{4\epsilon^4}{\hbar^2\rho^4 v^2}\sin^2\gamma_1(n00|z|m10)^4. \qquad \text{(V.33)}$$

To find a_1, still under the assumption $a_2 = a_3 = 0$:

$$|a_1|^2 + |b_1|^2 + |b_2|^2 + |b_3|^2 + |c_1|^2 = |a|^2 + |c|^2. \qquad \text{(V.34)}$$

and, if we suppose $|c_1|^2=|c|^2$, the change in $|a|^2$ may be determined from Eq. (V.34) as

$$\Delta|a|^2=|a_1|^2-|a|^2=-(|b_1|^2+|b_2|^2+|b_3|^2)$$

$$=-|a|^2\frac{\epsilon^4}{m^2\omega_{nm}^2}f_{nm}^2\frac{1}{\rho^4v^2}\sin^2\gamma_1 \qquad (V.35)$$

where

$$f_{nm}=(2m\omega_{nm}/h)(n00|z|m10)^2,$$

the oscillator strength of the mn transition.

We earlier set the condition $\mathcal{E}=E$ for determining ρ_0. Similarly, we now establish the criteria for ρ_0 as $\Delta|a|^2=|a|^2$. From Eq. (V.35) ρ_0 may then be defined as

$$\rho_0=\left[\frac{\epsilon^2}{m\omega_{nm}}f_{nm}\frac{1}{v}\right]^{\frac{1}{2}}. \qquad (V.36)$$

If there are N atoms per unit of volume of our gas then surely $2\pi\rho d\rho vN$ of them appear per unit of time lying a distance between ρ and $\rho+d\rho$ from the emitter. Furssow and Wlassow asserted this to mean that the total change of the probability per unit of time is the result of averaging over ρ:

$$\frac{d}{dt}|a|^2=-\pi|a|^2\frac{\epsilon^2}{m\omega_{nm}}f_{nm}N\sin^2\gamma_1. \qquad (V.37)$$

If Eq. (V.37) is averaged over γ, and the resulting equation solved for $|a|^2$, the quantum-mechanical analog of Eq. (V.26a) is obtained:

$$|a|^2=|a_0|^2e^{-\gamma t} \qquad (V.38)$$

where

$$\gamma=\frac{2\pi}{3}\frac{\epsilon^2}{m\omega_{nm}}f_{nm}N.$$

Thus the state lifetime has been decreased by the energy exchange as predicted, and this decrease is dependent on the gas density.

If we simply consider the atom-field interaction, then the state growth equations are given by

$$i\hbar\dot{a}_{m00\cdots0_s0\cdots}(t)$$

$$=\mathcal{H}_{m00\cdots0_s0\cdots}^{n00\cdots1_s0\cdots}a_{n00\cdots}(t)e^{i(\omega_{nm}-\omega_s)t} \qquad (V.39)$$

of solution

$$a_{n00\cdots1_s0\cdots}(t)$$

$$=\mathcal{H}_{m00\cdots0_s0\cdots}^{n00\cdots1_s0\cdots}\frac{e^{i(\omega_s-\omega_{nm})t-\gamma t/2}-1}{-\hbar[\omega_s-\omega_{nm}+i(\gamma/2)]} \qquad (V.40)$$

and we may recall that

$$I(\omega)=\text{const}|a_{r00\cdots1_s0\cdots}(\infty)|^2$$

$$=\frac{\text{const}}{(\omega-\omega_{nm})^2+(\gamma/2)^2}. \qquad (V.41)$$

From Eq. (V.41) it is apparent that our half-width is given by the following:

$$\delta_k=\gamma=\frac{2\pi}{3}\frac{\epsilon^2}{m\omega_{nm}}f_{nm}N. \qquad (V.42)$$

It can be seen that Eq. (V.42) yields the same results for the half-width as Eq. (V.28) which latter equation was found for the case of distant collisions.

For the case of near collisions let us utilize the Heisenberg uncertainty principle in the form

$$\Delta E\Delta t=\hbar$$

where we shall now assume ΔE to be the width of the state and Δt to be the mean lifetime of the state.

Let us suppose that the life of a state is terminated by a collision so that $\Delta t=\tau$ where τ is the mean time between collisions. This means that the width of the state will now be

$$\Delta E=\frac{\hbar}{\Delta t}=\frac{\hbar}{\tau}=2\hbar\pi\frac{\epsilon^2}{m\omega_0}f_{nm}N$$

or

$$\delta_I=\Delta\omega=\frac{\Delta E}{\hbar}=2\pi\frac{\epsilon^2}{m\omega_0}f_{nm}N. \qquad (V.43)$$

Finally we obtain for the half-width

$$\delta=\delta_k+\delta_I=\frac{8\pi}{3}\frac{\epsilon^2}{m\omega_0}f_{nm}N \qquad (V.44)$$

which is the same as the classical result given by Eq. (V.30).

In the main Furssow and Wlassow considered their results to hold only in the case of low gas densities. We may question the Furssow-Wlassow procedure in which they work out the broadening effect of one like atom and then simply extend these results to the case of N similar atoms. Because of the procedure Furssow and Wlassow felt that their results were only applicable to low gas densities. In a later paper[20] on the subject they investigated the self-broadening of a spectral line in the case of high gas densities. They found it necessary to use a slightly different although comparable method of approach to which we now turn our attention.

H. High-Pressure Quantum Resonance Broadening

The previous theory was really predicated on a collision between two particles. Now if, as in the Schulz and Rompe experiments, we take $\omega_0=1.02\times10^{16}$ sec^{-1}, $f=1.3$, $T=6000°$, we will obtain a ρ of approximately 5.35×10^{-7}. Then with the pressure used by Schulz and Rompe $N=2.57\times10^{19}$ cm^{-3}. This means that in the "sphere of action" we shall have $(4/3)\pi\rho^3N=17$ molecules. Thus, the approximation of a two-particle collision is hardly reasonable.

A modification of the first theory is certainly called for, on the basis of these considerations.

Let us begin by supposing the gas density sufficiently high that we can expect a large number n of atoms within the "sphere of action." If we allow this system of $n+1$ atoms to possess two levels, the resonance level and the ground state, methods of degenerate perturbation theory lead to the state growth equations:

$$i\hbar\dot{a}=\sum_{k=1}^{n}(a|U|b_k)b_k \quad \text{(V.45a)}$$

$$i\hbar\dot{b}_k=(b_k|U|a)a+\sum_{k'}(b_k|U|b_{k'})b_{k'}. \quad \text{(V.45b)}$$

Again under the initial conditions stipulating that atom 0 is initially excited we obtain $a(t)=a(0)=1$ and $b(0)=0$. Subsequent to the time $t=0$ we can expect the transfer of energy from the initially excited to the unexcited atom to proceed as governed by the matrix element $(a|U|b_k)$. In addition Eq. (V.45) tells us that there will be a secondary transfer of energy among the initially unexcited atoms as governed by the matrix element $(b_k|U|b_{k'})$. We now introduce the approximation that the secondary energy transfer process as governed by $(b_k|U|b_{k'})$ can be disregarded and suppose U to be time independent. This latter is nothing more nor less than saying that the thermal motion of the atom is so small as to be considered negligible during the energy transfer process. As we shall see, at high pressure for the resonance level this is a reasonable approximation. We now substitute for $\dot{b}_k$ from the thus modified Eq. (V.45b) into a differentiated Eq. (V.45a). This yields

$$-\hbar^2\ddot{a}=a\sum_{k=1}^{n}(a|U|b_k)^2. \quad \text{(V.46)}$$

If we now ignore the angular dependence of the dipole interaction there results

$$\ddot{a}+p^2a=0 \quad \text{(V.47a)}$$

$$p^2=\sum_{k=1}^{n}\frac{\alpha}{R_k^6};\quad \alpha=\frac{\epsilon^4f^2}{4m^2\omega_0^2}. \quad \text{(V.47b)}$$

A solution of Eq. (V.47a) which satisfies the initial condition $a(0)=1$ is

$$a=\cos pt. \quad \text{(V.48)}$$

A collision has occurred when $a=0$ so that

$$\tau_0=\pi/2p. \quad \text{(V.49)}$$

We assume the step function

$$\begin{aligned} a&=1 \quad \text{for} \quad 0\leqslant t\leqslant \tau_0 \\ a&=0 \quad \text{for} \quad 0<t>\tau_0 \end{aligned} \quad \text{(V.50)}$$

for a so that the Fourier integral for a is

$$a=\int_{-\infty}^{+\infty}g(\omega)e^{i\omega t}d\omega'$$
$$g(\omega)=\frac{1}{2\pi}\frac{1-e^{-i\omega\tau}}{i\omega}. \quad \text{(V.51)}$$

We average the absolute square of this expression over all collision times in order to obtain the intensity distribution in the level:

$$I(\omega)=\int_0^\infty|g(\omega)|^2w(\tau_0)d\tau_0. \quad \text{(V.52)}$$

Let us utilize the expression

$$\nu=\frac{\pi^2}{4\tau_0^2}=p^2 \quad \text{(V.53)}$$

for a change in variable in Eq. (V.52) as follows:

$$w(\tau_0)d\tau_0=\frac{\pi w(\pi/2\sqrt{\nu})}{4\sqrt{\nu^3}}d\nu=-I'(\nu)d\nu. \quad \text{(V.54)}$$

In order to find $I'(\nu)d\nu$ we suppose a random distribution of spatial coordinates to exist. Let us establish a configuration space of $3n$ dimensions whose volume is $\mathfrak{U}=V^n$ where V is the volume of our gas. Now if we temporarily disregard the interatomic forces, we can expect equal probability for the occupation of any portion of this space. This means that the followlng relation will hold:

$$I'(\nu)d\nu=\frac{\Delta\mathfrak{U}}{\mathfrak{U}}, \quad \text{(V.55)}$$

where $\Delta\mathfrak{U}$ represents that portion of configuration space for which our ν lies between ν and $\nu+d\nu$. This value of ν will, of course, depend on the distribution of the R_k in Eq. (V.47b). Thus

$$\Delta\mathfrak{U}=(4\pi)^n\int\cdots\int R_1^2R_2^2\cdots R_n^2dR_1dR_2\cdots dR_n. \quad \text{(V.56)}$$

The analogy to Margenau's equation (III.4) is apparent. Our solution then is given by Eq. (III.18) as

$$I'(\nu)=\tfrac{2}{3}\pi N\sqrt{\alpha}\frac{\exp\left(-\dfrac{4\pi^3\alpha N^2}{9\nu}\right)}{\sqrt{\nu^3}}. \quad \text{(V.57)}$$

We may now transform back to our variable τ_0 by utilizing Eq. (V.54). We obtain

$$w(\tau_0)=\frac{2\gamma}{\sqrt{\pi}}\exp(-\gamma^2\tau_0^2) \quad \text{(V.58)}$$

$$\gamma=(4/3)N(\pi\alpha)^{\frac{1}{2}}=\tfrac{2}{3}\sqrt{\pi}\frac{f\epsilon^2}{m\omega_0}N. \quad \text{(V.59)}$$

Equations (V.49) and (V.58) may now be substituted into Eq. (V.52) and the results integrated to obtain

$$I(\omega)=\frac{1}{2\pi^2}\frac{1-\exp[-(\omega/2\gamma)^2]}{\omega^2} \tag{V.60}$$

as the distribution of energies in the resonance level. For the breadth we obtain

$$\delta=2.54\gamma=3\frac{f\epsilon^2}{m\omega_0}N. \tag{V.61}$$

In order to find the spectral line width, it would appear reasonable to assume that we must needs have a knowledge of not only the resonance level width but also the width of the level from which the radiating transition originates.††† The phrase "from which" leads us to the next consideration.

Furssow and Wlassow felt that these results should only be applied to transitions proceeding from upper levels to the resonance level, that is, these results should not be applied, to the transition from the resonance level to the ground state. A short consideration renders this assertion plausible.

Let us suppose that the initially excited atom undergoes a transition from the resonance level to the ground state with the accompanying emission of radiation. This emitted radiation may be absorbed by one of the unexcited atoms where absorption would not be possible were this radiation the result of a transition to some level above the ground state. This process, which should not be confused with the transfer of energy without accompanying radiation has certainly not been considered in the theory, and, since it can be expected to have some effect on the line broadening, this theory cannot be expected to hold in such cases.

The justification for the assumption of fixed atoms, that is, $d/dt(a|U|b_k)=0$ which led to Eq. (V.46) appears worthy of note here. Let us consider the Rompe and Schulz case.[75] The pressure is 80 atmospheres and $T=2800°$. The experimental width was found to be $\delta=7\times10^{12}$ sec^{-1} so that the mean life of the resonance level is $\tau_0=1/\delta=1.4\times10^{-13}$ sec. The mean relative velocity is $\langle v\rangle=1.3\times10^5$ cm/sec, which leads us to the conclusion that during the mean level life we may expect our atoms to move a distance $\langle v\rangle\tau_0=1.8\times10^{-8}$ cm, or the order of their own diameter. Thus, the approximation appears a reasonable one.

Furssow and Wlassow's high pressure theory yields results in reasonable agreement with those of Schulz and Rompe.

I. Comparison of Some Experimental Results

Three of the equations for the half-width which we have obtained are

Weisskopf: $\delta=\frac{\epsilon^2}{m\omega_0}f_{nm}N$

Margenau and Watson: $\delta=\frac{\pi}{3}\frac{\epsilon^2}{m\omega_0}f_{nm}N$

Furssow and Wlassow: $\delta=\frac{4}{3}\frac{\epsilon^2}{m\omega_0}f_{nm}N.$

These three equations have been compared with the experimental results, for rubidium by Chen,[77] for sodium by Watanabe,[89] and for caesium by Gregory.[23]

Other than demonstrating the fact that none of the theories propounded in explanation of the resonance broadening phenomenon is precisely correct these experimental results indicated that the combined resonance and, say, van der Waals forces should be investigated. Such an investigation might provide the explanation of the observed line shape variations with radiating transition as well as with pressure.

VI. MOLECULAR BROADENING

A. Early Studies of Polyatomic Molecular Broadening

In beginning our considerations of the last chapter we remarked that those studies which had gone before had not distinguished between broadeners of a type different from the emitter and broadeners of the same type as the emitter. In this same vein we might now remark that, for all practical purposes, we have not differentiated between atoms and polyatomic molecules, for the broadening phenomena to which we have so far devoted some little attention are equally as applicable to the latter as to the former. It is really by the appearance of rotation and vibration spectra that we may differentiate between the polyatomic and monatomic molecule so that by such a statement we mean that these previous broadening considerations are properly applicable to this type spectra as well as to electronic spectra. It was originally felt by such authors as Kussmann[44] and Lasareff[45] and Grasse[22] that these theories were not only applicable but also sufficient. The intimation that some different approach to the broadening of rotation-vibration lines was in order appears to have arisen first in the work of Herzberg and Spinks.[27] These investigators found a decrease of the line width with higher values of the rotational quantum number J in the near infrared spectra of HCN. This would surely indicate that rotation-vibration spectra should be treated differently than electronic spectra, but this clear indication was rather clouded when the work of Cornell and Watson[11] and Herzberg, Spinks, and Watson[27] failed to verify this variation with J. The hint had been given however, and it was certainly apparent that the possibility of J width dependence and other unique molecular phenomena should be theoretically investigated.

††† The complexity of computing the width of a higher level may be inferred from the preceding work.

B. The Directional Effect and Rotational Resonance

In studying broadening phenomena one usually looks first to the molecular or atomic interaction to which we can attribute responsibility for the broadening in question. In molecular broadening an interaction which plays quite a major role is the directional effect, a special case of which is the rotational resonance effect, to which we now turn our attention.

Let us specify the dipole moments of two rotators by μ_1 and μ_2 and their separation by R. In this case then, the dipole-dipole interaction leads to the interaction energy

$$\langle E^{(2)}\rangle = -\frac{2}{3}\frac{\mu_1^2\mu_2^2}{R^6}\frac{1}{(2J_1+1)(2J_2+1)} \times\left\{\left[\frac{2A_1}{J_2+1}+\frac{2A_2}{J_1+1}\right]^{-1}+\left[\frac{2A_1}{J_2}-\frac{2A_2}{J_1+1}\right]^{-1} -\left[\frac{2A_1}{J_2+1}-\frac{2A_2}{J_1}\right]^{-1}+\left[\frac{2A_1}{J_2}+\frac{2A_2}{J_1}\right]^{-1}\right\} \quad \text{(VI.1)}$$

where the J_i refer to the lower state, and $A_i=\hbar^2/2I_i$.

If the two molecules are identical so that $A_1=A_2$ and $\mu_1=\mu_2$ Eq. (VI.1) becomes

$$\langle E^{(2)}\rangle = -\frac{2}{3}\frac{\mu^4 I}{\hbar^2 R^6} \times\frac{J_1(J_1+1)+J_2(J_2+1)}{(J_1+J_2)(J_1+J_2+2)(J_1-J_2-1)(J_1-J_2+1)}. \quad \text{(VI.2)}$$

A special case arises when the broadener is in the ground state, a case for which we shall find specific use:

$$\langle E^{(2)}\rangle = -\frac{2}{3}\frac{\mu_1^2}{R^6}\frac{1}{2J_2''+1}\left\{\sum{}'|(00|\epsilon r|J_1'M_1')|^2 \times\left(\frac{J_2''+1}{E_{J_1'}-E_0-2(J_2''+1)A} +\frac{J_2''}{E_{J_1'}-E_0+2J_2''A}\right)\right\}. \quad \text{(VI.3)}$$

In passing, let us note a corresponding expression for the dipole-dipole interaction of two symmetric top molecules possessed of dipoles as given by Carroll[7]:

$$\langle E^{(2)}\rangle = -\frac{2}{3}\frac{\mu^4 I_A}{R^6\hbar^2}\left\{\frac{(J_1+1)^2-K_1^2}{(J_1+1)(2J_1+1)}\left[\frac{(J_2+1)^2-K_2^2}{(J_2+1)(2J_2+1)(J_1+J_2+2)}+\frac{K_2^2}{J_2(J_2+1)(J_1+1)}+\frac{J_2^2-K_2^2}{J_2(2J_2+1)(J_1-J_2+1)}\right]\right.$$
$$\left.+\frac{K_1^2K_2^2}{J_1(J_1+1)J_2^2(J_2+1)^2}+\frac{J_1^2-K_1^2}{J_1(2J_1+1)}\left[\frac{(J_2+1)^2-K_2^2}{(J_2+1)(2J_2+1)(J_2-J_1+1)}-\frac{K_2^2}{J_1J_2(J_2+1)}-\frac{J_2^2-K_2^2}{J_2(2J_2+1)(J_1+J_2)}\right]\right\}. \quad \text{(VI.4)}$$

A consideration of Eq. (VI.2) tells us that a rotational resonance condition sets in for identical rotators for $|J_1-J_2|=1$. Further, Eq. (VI.2) indicates that for dissimilar dipoles an accidental rotational resonance condition comes about when $J_2A_2=(J_1+1)A_1$ or $(J_2+1)A_2=J_1A_1$. For these cases a first-order perturbation energy exists and must be evaluated.

The method of evaluation which is applicable here is the variational perturbation treatment. In carrying out this treatment one obtains a determinant of the form, say, Eq. (V.7) whose order is, of course, dependent on the J values involved. This means that no solution for the general case (any J value, as, for example, is given by Eq. (VI.1)) is obtainable since no such general solution to the determinant is available. The root-mean-square value of the first-order energy perturbation is the same for all determinants, however, and it may be obtained as

$$\langle[\epsilon^{(1)}]^2\rangle^{\frac{1}{2}} = \left(\frac{2}{3}\right)^{\frac{1}{2}}\frac{J+1}{[(2J+1)(2J+3)]^{\frac{1}{2}}}\frac{\mu^2}{R^3}. \quad \text{(VI.5)}$$

As an example of the specific calculation, if we let $J_1=0$, $J_2=1$, the perturbation amounts to

$$E^{(1)} = \pm\frac{1}{3}\frac{\mu^2}{R^3}, \quad \pm\frac{2}{3}\frac{\mu^2}{R^3}. \quad \text{(VI.6)}$$

A further consideration of Eq. (VI.2) suffices to show that $J_1=J_2=0$ also leads to a rotational resonance condition, but this case corresponds to the precise solution:

$$E^{(1)} = \frac{\hbar^2}{I}\left[1-\left(1+\frac{2\mu^4I^2}{3\hbar^4R^6}\right)^{\frac{1}{2}}\right]. \quad \text{(VI.7)}$$

Equation (VI.7) tells us the first order interaction energy for two like dipole linear rotators for the case $J_1=J_2=0$. For large R, the radical in this equation may be expanded with the result

$$E^{(1)} \doteq -\frac{\mu^4 I}{3\hbar^2 R^6} \quad \text{(VI.8a)}$$

and for small R Eq. (VI.7) becomes

$$E^{(1)} \doteq -\left(\frac{2}{3}\right)^{\frac{1}{2}} \frac{\mu^2}{R^3}. \quad \text{(VI.8b)}$$

Let us now sketch a resonance development of London in just enough detail so that the results which we shall later utilize will be somewhat intelligible. This author used what we might call an order-of-magnitude technique to arrive at the secular determinant

$$|U_{\Upsilon_i\Upsilon_j}+(E_{\Upsilon_i}^{(0)}-E_{\Upsilon_k}^{(0)}-E_k^{(1)})\delta_{\Upsilon_i\Upsilon_j}|=0. \quad \text{(VI.9)}$$

We shall be able to define the new symbols $E_{\Upsilon_i}^{(0)}$ and $E_{\Upsilon_k}^{(0)}$ after another step in the development.

London considered only the eigenfunction $\psi_{J_1}^{M_1}\psi_{J_2}^{M_2}$ and the thirteen eigenfunctions with which this eigenfunction may combine, as always, under the influence of U, thus rendering his treatment an approximate one.

Under these restrictions then, we shall utilize the eigenfunctions:

$$\begin{aligned}
&u_0=\psi_{J_1}^{M_1}\psi_{J_2}^{M_2}\\
&\left.\begin{aligned}u_1&=\psi_{J_1+1}^{M_1}\psi_{J_2+1}^{M_2}\\ u_2&=\psi_{J_1+1}^{M_1+1}\psi_{J_2+1}^{M_2-1}\\ u_3&=\psi_{J_1+1}^{M_1-1}\psi_{J_2+1}^{M_2+1}\end{aligned}\right\}\Upsilon_1\\
&\left.\begin{aligned}u_4&=\psi_{J_1-1}^{M_1}\psi_{J_2-1}^{M_2}\\ u_5&=\psi_{J_1-1}^{M_1+1}\psi_{J_2-1}^{M_2-1}\\ u_6&=\psi_{J_1-1}^{M_1-1}\psi_{J_2-1}^{M_2+1}\end{aligned}\right\}\Upsilon_4\\
&\left.\begin{aligned}u_7&=\psi_{J_1+1}^{M_1}\psi_{J_2-1}^{M_2}\\ u_8&=\psi_{J_1+1}^{M_1+1}\psi_{J_2-1}^{M_2-1}\\ u_9&=\psi_{J_1+1}^{M_1-1}\psi_{J_2-1}^{M_2+1}\end{aligned}\right\}\Upsilon_7\\
&\left.\begin{aligned}u_{10}&=\psi_{J_1-1}^{M_1}\psi_{J_2+1}^{M_2}\\ u_{11}&=\psi_{J_1-1}^{M_1+1}\psi_{J_2+1}^{M_2-1}\\ u_{12}&=\psi_{J_1-1}^{M_1-1}\psi_{J_2+1}^{M_2+1}\end{aligned}\right\}\Upsilon_{10}
\end{aligned} \quad \text{(VI.10)}$$

from which we may define $E_{\Upsilon_i}^{(0)}$ and $E_{\Upsilon_k}^{(0)}$.

$E_{\Upsilon_k}^{(0)}$ is in all cases the zeroth-order energies going with u_0. $E_{\Upsilon_i}^{(0)}$ are the zeroth-order energies going with $u_1, u_2, \cdots, u_{12}$. We may note that the energies associated with u_1, u_2, u_3 are the same as are those associated with u_4, u_5, u_6 and so forth. Thus

$$E_1=E_{\Upsilon_1}^{(0)}-E_{\Upsilon_0}^{(0)}=(J_1+J_2+2)\frac{\hbar^2}{I}, \text{ etc.} \quad \text{(VI.11)}$$

A secular determinant results. In addition, we let $U_{0i}=U_i$ and from this secular determinant, the equation

$$R^6E^{(1)}=\frac{a_1^2}{E^{(1)}-E_1}+\frac{a_2^2}{E^{(1)}-E_2}+\frac{a_3^2}{E^{(1)}-E_3}+\frac{a_4^2}{E^{(1)}-E_4} \quad \text{(VI.12)}$$

follows, where

$$\frac{a_i^2}{R^6}=U_{3i}^2+U_{3i-1}^2+U_{3i-2}^2. \quad \text{(VI.13)}$$

For the resonance case ($J_1>0$, $J_2>0$, $J_1=J_2+1$) Eq. (VI.12) becomes

$$R^6E^{(1)}=\frac{a_1^2}{E^{(1)}-E_1}+\frac{a_2^2}{E^{(1)}-E_2}+\frac{a_3^2}{E^{(1)}-E_3}+\frac{a_4^2}{E^{(1)}}. \quad \text{(VI.14a)}$$

For large R and $|E^{(1)}|\ll|E_i|$:

$$R^6(E^{(1)})^2+\left(\frac{a_1^2}{E_1}+\frac{a_2^2}{E_2}+\frac{a_3^2}{E_3}\right)E^{(1)}-a_4^2=0 \quad \text{(VI.14b)}$$

whose solution is

$$E^{(1)}\doteq-\tfrac{1}{2}\left(\frac{a_1^2}{E_1}+\frac{a_2^2}{E_2}+\frac{a_3^2}{E_3}\right)\frac{1}{R^6}\pm\frac{a_4}{R^3}. \quad \text{(VI.15)}$$

Equation (VI.15) gives the first-order perturbation energy for the rotational resonance case. We note the comparison with Eq. (VI.6).

C. Broadening by Molecules with No Permanent Poles

Margenau[59] initiated the first attack on the theory of molecular broadening in 1936. He chose to investigate the case of broadening by foreign gases which possess no permanent poles since Watson and Hull[91] had previously obtained experimental data on this case. These authors had investigated the broadening of AlH by H_2 and concluded that all rotation-vibration lines associated with a particular electronic transition would be broadened by the same amount.

From Eq. (VI.3) we may obtain the perturbation on the energy of the absorber averaged over all orientations of the absorber as

$$\langle E^{(2)}\rangle=-\frac{2}{3R^6}\sum_{\Lambda\mathfrak{H}}{}'\frac{|(\epsilon r)_{0\Lambda}|^2|(U)_{k\mathfrak{H}}|^2}{E_\Lambda-E_0+E_{\mathfrak{H}}-E_k}, \quad \text{(VI.16)}$$

where $\mathfrak{H}$ and Λ refer to absorber and broadener, respectively.

We have stipulated that E_0 is the lowest energy of the perturber. This means that $E_\Lambda-E_0$ will always be positive. In addition we have not allowed this perturber a permanent dipole moment. This means that no pure rotational transitions, which amount to the lowest energy transitions, are possible. Thus the vibrational transitions are the lowest energy allowed.

On the other hand, $E(\mathfrak{H})-E_k$ may be positive or negative. Consequently $\langle E^{(2)}\rangle$ may be greater than or less than zero so that the spectral line may exhibit either a violet or a red asymmetry.

Now where we simply have dealt with a set of relatively widely spaced electronic levels in our monatomic considerations, we now also deal with the closely spaced superposed rotation-vibration levels. When we are considering the visible and ultraviolet portion of the spectrum we may, as an approximation, consider only

the electronic contributions to $E(\mathfrak{H})-E_k$—they surely constitute the largest ones—and merely let the rotation-vibration levels superposed upon them amount to degeneracies accounted for by a summation in Eq. (VI.20). It follows that the result will closely correspond to that for monatomic molecules.

In order to consider the possible effect of molecular rotation let us approximate our absorber by a rigid rotator of dipole moment μ. Now I is usually of the order of 10^{-40}, and, with h of the order of 10^{-27} we see that A will be small compared to $E_\Lambda - E_0$ when the perturber has no permanent dipole. When A is neglected Eq. (VI.3) becomes

$$\langle E^{(2)}\rangle = -\frac{2}{3}\frac{\mu^2}{R^6}\sum_\Lambda{}' \frac{|(\epsilon r)_{0\Lambda}|^2}{E_\Lambda - E_0} = -\frac{\alpha\mu^2}{R^6}. \quad \text{(VI.17)}$$

If the dipole moment $\mu \doteq 10^{-18}$, the perturbation given by Eq. (VI.17) is about 1/20 of that for the electronic case. Now let us write Eq. (VI.3) as

$$\langle E_J{}^{(2)}\rangle = -\frac{\alpha\mu^2}{R^6}\left[1 + \frac{2A}{(E_\Lambda - E_0)} + \frac{4(J^2+J+1)A^2}{(E_\Lambda - E_0)^2} + \cdots\right]. \quad \text{(VI.18)}$$

It is apparent from Eq. (VI.22) that the portion of $\langle E_J{}^{(2)}\rangle$ which depends on J is very small due to the appearance of A^2 in the numerator and $(E_\Lambda - E_0)^2$ in the denominator involved.

Margenau carried out his considerations of the effect of molecular vibrations in an analogous manner. In this case we obtain

$$\langle E_v{}^{(2)}\rangle = 3\frac{\epsilon^2 E_v}{m\omega^2 R^6}\alpha. \quad \text{(VI.19)}$$

On the basis of these investigations, we may conclude that within the limits of experimental error all rotation-vibration lines associated with the same electronic transition should show broadening and shift of about the same degree when the broadening agent is a nonpolar molecule. In addition these lines should be affected by foreign perturbers almost in the manner of the corresponding monatomic lines under similar circumstances.

As Watson[90] has noted "dispersion forces"—the type considered—also appear to predominate in the self-broadening of nonpolar molecular lines. Were the normal resonance forces between like monatomic molecules present, we would not normally expect this to be the case, but the closely spaced rotational levels of the polyatomic molecule render this predominance reasonable. As we have noted, the rotational levels specified by, say, just J are comparatively closely spaced energywise. As a consequence a Maxwell-Boltzmann temperature distribution of molecules over these levels will not lead to the preferential population of the monatomic case, and they will be more equally divided among many levels. On the other hand, the allowed J change in transition is generally limited to ± 1 or 0 so that the probability of energy exchange with the resultant degeneracy is severely restricted. In toto, then, this resonance effect would not appear to be primarily responsible for self-broadening, and we may safely look to these dispersion forces as broadening agents for nonpolar molecules. This conclusion had been verified for the case of methane by Childs,[8] Dennison and Ingram,[14] and Vedder and Mecke,[85] for atmospheric pressure and above, and for the case of acetylene by Hedfield and Mecke,[24] Herzberg and Spinks,[25] Lochte-Holtgreven and Eastwood,[52] and McKellar and Bradley[65] for atmospheric pressure.

We have broadened molecular lines by nonpolar molecules, and it would appear a logical next step to broaden these lines by polar molecules.

D. Qualitative Explanations of Broadening by Dipole Molecules

It is to the directional effect of Sec. B that we may look for an explanation of the relatively large, J-dependent broadening effects due to dipole broadeners.

To begin with we must hypothesize unpolarizable (rigid) dipoles, and we recall from Sec. B that (1) the intermolecular forces are relatively weak except when the near resonance condition, $|J_1 - J_2| < 2$ or 3, set in and (2) the lowest rotational state is most strongly influenced by other molecules in the lowest of neighboring states.

The lower state J_1'' we take as other than the lowest rotational state so that we can concern ourselves with (1). As has been mentioned, the rotational levels lie sufficiently close together so that a thermal distribution over them may be expected. Suppose J_1'' to designate a level near the maximum of this thermal distribution. Then a much greater number of perturbers can be expected to fulfill the condition $|J_1'' - J_2''| < 2$ or 3 than would be the case were the J_1'' level to be found toward the wings of this thermal distribution. Since the upper level will also enter into the line broadening one cannot say more as yet than that those most intense lines arising from J levels near the maximum of the thermal distribution should be broadened more than the remainder of the lines of the band.

In regard to (2) this indicates that a greater broadening of the line arising from the lowest rotational level is to be expected, but only under certain conditions. At normal temperatures, the thermal distribution should not populate levels lying near $J_1''=0$ very highly so that the effect may not be very pronounced. As the temperature is lowered, however, these low level populations will increase and we should expect to see an increase in the broadening of the line with $J_1''=0$ over neighboring spectral lines. It might be noted also that due to "greater flexibility downward" a violet asymmetry in this line may be expected.

We now devote ourselves to Margenau and Warren's more quantitative consideration of the broadening effected by dipole interactions.

E. Statistical Broadening by Symmetrical Top Dipole Molecules

For our two interacting molecules we choose two identical symmetrical rotators each possessed of a permanent electric dipole moment μ oriented along the molecular figure axis. The interaction potential between the two rotators is again given by Eq. (VI.1) where now $\mu_1=\mu_2=\mu$. The eigenfunctions for this system are the familiar product functions for sufficient separation R. U is not dependent on the Eulerian angle χ so that $K_1''=K_1'$ and $K_2''=K_2'$ under the aegis of U, and the matrix of U is diagonal in K_1 and K_2 as a consequence.

After a reasonable amount of mathematical manipulation one obtains

$$\begin{aligned}(M_1''M_2''|U|M_1'M_2') &= \frac{\mu^2}{R^3}\frac{K_1K_2}{J_1(J_1+1)J_2(J_2+1)} \\ &\times\{-2M_1''M_2''\delta(M_1'M_1'')\delta(M_2'M_2'') \\ &\pm\tfrac{1}{2}[(J_1-M_1''+1)(J_1+M_1'')(J_2-M_2'') \\ &\times(J_2+M_2''+1)]^{\frac{1}{2}}\delta(M_1', M_1''-1)\delta(M_2', M_2''+1) \\ &\pm\tfrac{1}{2}[(J_1+M_1''-1)(J_1-M_1'')(J_2-M_2''+1) \\ &\times(J_2+M_2'')]^{\frac{1}{2}}\delta(M_1', M_1''+1) \\ &\times\delta(M_2', M_2''-1)\}. \qquad (VI.20)\end{aligned}$$

From Eq. (VI.20) it is apparent that the first-order perturbation energy does not disappear. The secular determinant breaks into blocks, with identical blocks symmetrical about the secondary diagonal. Thus every root of this determinant will be a double root so that general equation describing the level splitting due to this interaction is not to be found, but an idea as to the maximum splitting may be obtained, and the example of a special case may be presented.

We may write down from Eq. (VI.20) the double root which lies at either end of the principal diagonal and where $J_1=M_1$ and $J_2=M_2$.

$$E^{(1)}=\mp\frac{2\mu^2}{R^3}\frac{K_1K_2}{(J_1+1)(J_2+1)}. \qquad (VI.21)$$

Figure 16 illustrates the level splitting with the rotator separation R and the magnetic quantum number M for the special case $J_1=K_1=2$, $J_2'=K_2'=1$ for the upper state and $J_1=K_1=1$, $J_2''=K_2''=1$ for the lower state.

Let us finally remark that since $K=0$ for a spherical top rotator, the matrix elements of U in Eq. (VI.20) would all be zero so that the effect considered would not take place.

Insofar as second-order effects are concerned, we certainly are aware that they will depend on the inverse sixth power of the molecular separation. In addition Margenau and Warren felt that for finding those regions in which the first order perturbations predominate, London's second-order results for diatomic molecules should be sufficient. Under these assumptions a limiting range of about $7A$ for the predominance of the forces which we have considered was imposed by these authors.

The root-mean-square energy perturbation may be obtained as

$$\langle[E^{(1)}]^2\rangle^{\frac{1}{2}}=\left(\frac{2}{3}\right)^{\frac{1}{2}}\frac{\mu^2}{R^3}\frac{K_1K_2}{[J_1(J_1+1)J_2(J_2+1)]^{\frac{1}{2}}}. \qquad (VI.22)$$

Now surely a reasonable approximation to the half-width arising from these considerations may be expected if we equate a to

$$\left(\frac{2}{3}\right)^{\frac{1}{2}}\mu^2\frac{|K_1K_2|}{[J_1(J_1+1)J_2(J_2+1)]^{\frac{1}{2}}}$$

in order to obtain the half-width

$$\delta=\frac{4}{3}\pi^2\frac{Na}{h}=2\left(\frac{2}{3}\right)^{\frac{1}{2}}\frac{\pi^2\mu^2|K_1K_2|N}{h[J_1(J_1+1)J_2(J_2+1)]^{\frac{1}{2}}} \qquad (VI.23)$$

for the statistical result.

It is true, of course, that we have here found the level width while we are really desirous of the spectral line width. The level here $(J_1K_1J_2K_2)$ would have to be

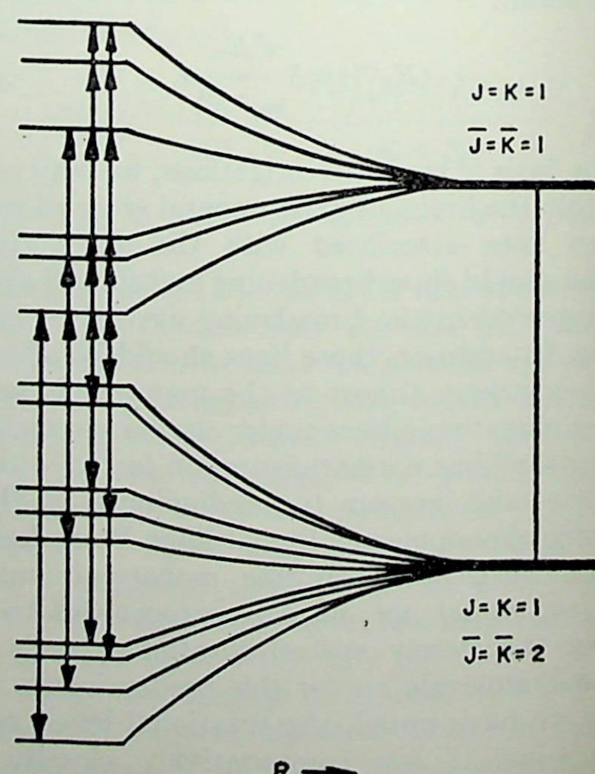


FIG. 16. The Rotational levels of the symmetric top rotator are split due to the dipole interaction. (After Margenau and Warren.[63])

considered in conjunction with another possible level $(J_1 \pm 1, 0; K_1 \pm 1, 0; J_2K_2)$. One may perhaps consider the K_1 and J_1 appearing in Eq. (VI.23) as the mean of the quantum numbers for the two states.

We might conclude by noting that the effect as given by Eq. (VI.23) is around 20 times as weak as the self-broadening effect in monatomic resonance lines.

The results of this theory were born out order of magnitude wise by Cornell's results.[10]

F. Broadening by the Linear Dipole Molecule HCN

The first attempt at a rigorous theoretical interpretation of the self-broadening of the absorption lines of a linear dipole molecule was carried out by Lindholm,[48] for the HCN molecule. In essence Lindholm used the energy perturbations due to the directional and resonance effects within the framework of the simple interruption theory to determine the broadening of the HCN lines in a manner which we now consider.

According to the simple interruption theory the half-width of the lines for a heterogeneous gas containing N_1, molecules with optical collision diameters ρ_1, N_2 with diameters ρ_2, etc.

$$\delta = \langle v \rangle \sum_i N_i \rho_i^2 \qquad \text{(VI.24)}$$

if all molecules are of the same mass, thus allowing a common $\langle v \rangle$.‡‡‡ Let us consider the necessity for introducing Eq. (VI.24).

From Lindholm's experimental results, it is quite apparent that a marked dependence on J is present in the line width. In the interruption theory, this can only come about through some variation in N and ρ with J_1 (the absorber J value) and J_2 (the perturber J value). Thus, we hypothesize a dependence of ρ on J_1 and J_2 which a little consideration immediately bears out. To begin with a frequency perturbation $\Delta\nu = b/R^6$ (which is brought about by the directional effect $|J_1 - J_2| \neq 1$) or $\Delta\nu = B/R^3$ (due to the resonance effect $|J_1 - J_2| = 1$) leads to the optical collision diameter

$$\rho = \left(\frac{3\pi^2}{4}\frac{b}{\langle v \rangle}\right)^{1/5} \qquad \text{(VI.25a)}$$

or

$$\rho = \left(\frac{4\pi B}{\langle v \rangle}\right)^{1/2}, \qquad \text{(VI.25b)}$$

respectively.

This then is at least a part of the general manner in which ρ may depend on J_1 and J_2, and let us now specificize this into usefulness. Lindholm dealt only with the P branches $(\Delta J = -1)$ of the two HCN bands which he considered. Insofar as the vibrational quantum numbers involved are concerned $v_1'' = v_2'' = 0$. We now let

$$\Delta = J_2'' - J_1'' \qquad \text{(VI.26a)}$$

so that

$$J_2'' - J_1' = J_2'' - (J_1'' - 1) = \Delta + 1 \qquad \text{(VI.26b)}$$

‡‡‡ Lindholm defines $\langle v \rangle$ as the mean relative velocity.

for the upper state since we consider only the P branch. Δ is now a convenient parameter for the determination of the type effect to be considered, and leads to three cases.

Case (1): Resonance effect in ground state and directional in upper $(\Delta = \pm 1)$.

Frequency perturbation:

$$\Delta\nu = \frac{b}{R^6} \pm \frac{B}{R^3} \qquad \text{(VI.27a)}$$

$$\int 2\pi\Delta\nu dt = \left|\frac{3\pi^2 b}{4\langle v \rangle}\frac{1}{\rho^5} \pm \frac{4\pi B}{\langle v \rangle}\frac{1}{\rho^2}\right| = 1. \qquad \text{(VI.27b)}$$

Case (2): Directional in ground and resonance in upper $(\Delta = 0, -2)$.

Same as Case (1).

Case (3): Directional effect in both states (all other Δ).

Frequency perturbation:

$$\Delta\nu = b/R^6 \qquad \text{(VI.28a)}$$

$$\rho = \left[\frac{3\pi^2}{4}\frac{b}{\langle v \rangle}\right]^{1/5}. \qquad \text{(VI.28b)}$$

In cases (1) and (2) ρ may be found as roots of Eq. (VI.27b), in particular there will be roots ρ_+ for the repulsive case (positive sign) and ρ_- for the attractive case (negative sign). In all cases of multiple roots ρ the highest valued root will be taken as significant. Further we shall suppose there to be equal amounts of resonance repulsion and attraction so that for ρ^2 from Eq. (VI.27b) we shall write $\rho_+^2 + \rho_-^2/2$. Now the ρ values furnished by the three possible sets of physical conditions may be substituted into Eq. (VI.24) with the result

$$\delta = 2.2(\langle v \rangle)^{3/5} \sum_i b_i^{3/5} N_i + \langle v \rangle \sum_j N_j \frac{\rho_{+j}^2 + \rho_{-j}^2}{2}. \qquad \text{(VI.29)}$$

Let us first consider the directional effect contributions. Lindholm approximated Eq. (VI.2) for this case by the expression

$$E_{J_1J_2}^{(2)} = \frac{1}{R^6}\frac{\mu^4 I}{3\hbar^2(\Delta^2 - 1)} \qquad \text{(VI.30)}$$

which he noted is asymptotic for large J, exact for $J_1 = J_2$, 86% of the correct value for $J_1 = 1$, $J_2 = 3$, and 92% of the correct value for $J_1 = 2$, $J_2 = 4$.

We obtain b for Eq. (VI.29) from Eq. (VI.30) as

$$b = \frac{R^6}{h}\left[E_{J_1'J_2'^{(2)}} - E_{J_1''J_2''^{(2)}}\right]$$

$$= \frac{4\pi^2\mu^4 I}{3h^3}\left[\frac{1}{(\Delta+1)^2 - 1} - \frac{1}{\Delta^2 - 1}\right] \qquad \text{(VI.31)}$$

according to Eqs. (VI.26).

In order to find N_i a Maxwell-Boltzmann distribution of the molecules over the rotational levels is assumed and the first term in Eq. (VI.29) now becomes

$$\delta_D = 2.2(\langle v\rangle)^{3/5}\left(\frac{4\mu^4\pi^2 I}{3h^3}\right)^{2/5} \times \frac{N.}{\sum'_{J_2=0}^{\infty} w_{J_2}} \sum_{J_2=0}^{\infty} w_{J_2}\left[\frac{1}{(\Delta+1)^2-1} - \frac{1}{\Delta^2-1}\right]^{2/5}, \quad \text{(VI.32)}$$

where the prime on the summation sign decrees the dropping of those J_2 values for which $\Delta = \pm 1, 0, -2$ (resonance effect).

Lindholm found it necessary to use the first twenty or thirty terms in the sum.

We first consider the case of both molecules in the ground state with $\Delta = \pm 1$. For the vibrational quantum numbers, the same Eq. (VI.19) is valid as it stands. London has evaluated a_1, a_2, and a_3 and Lindholm evaluated a_4 so that Eq. (VI.15) becomes

$$E^{(1)} = -\frac{\pi^2\mu^4 I}{6h^2}\frac{1}{R^6} \pm \frac{\mu^2}{2.52}\frac{1}{R^3}. \quad \text{(VI.33)}$$

Equation (VI.33) then gives the *first-order* interaction energy for this resonance condition, in the lower state (Case 1). Equations (VI.27) must now be solved for ρ. In Eq. (VI.27b) B is simply given by the coefficient of R^{-3} in Eq. (VI.33), but b presents a slightly different case. Since in the upper state, we have the directional effect taking place, b_{upper} will be given by the coefficient of R^{-6} in Eq. (VI.30). On the other hand, the resonance effect of the lower state furnishes b_{lower} which is the coefficient of the R^{-6} term in Eq. (VI.33). Finally

$$b = b_{\text{upper}} - b_{\text{lower}}. \quad \text{(VI.34)}$$

The values of B and b thus obtained may now be used in Eq. (VI.27b) to yield ρ values which when properly utilized in Eq. (VI.29) yield

$$\delta_R = \langle v\rangle \frac{N}{\sum_{J'=0}^{\infty} w_{J'}} \times [w_{J+1}7.71\times10^{-14} + w_{J-1}5.95\times10^{-14}]. \quad \text{(VI.35)}$$

Finally Case (2) remains to be investigated, and we introduce a vibrating rotator whose moments of inertia are dependent on the vibrational state. In correspondence with the u_0 of Eq. (VI.14) let us take

$$u_0' = \psi_{J+1}{}^{M_1}\psi_v(1)\psi_J{}^{M_2}\psi_{v'}(2)$$

and now there will be twenty-four additional eigenfunctions of this form which may combine with this one—we assume that v is the ground vibrational state and v' some resonating upper vibrational state. Corresponding to each of London's matrix elements $U_i = U_{0i}$ there are now two matrix elements. As an example, to U_1 there correspond

$$U_{01} = U_1' = \int \psi_{J+1}{}^{M_1}\psi_v(1) \times \psi_J{}^{M_2} U \psi_{J+2}{}^{M_1}\psi_v(1)\psi_{J+1}{}^{M_2}\psi_{v'}(2)d\sigma \quad \text{(VI.36a)}$$

$$U_1'' = \int \psi_{J+1}{}^{M_1}\psi_v(1)\psi_J{}^{M_2}\psi_{v'}(2) \times U\psi_{J+2}{}^{M_1}\psi_{v'}(1)\psi_{J+1}{}^{M_2}\psi_{v'}(2)d\sigma. \quad \text{(VI.36b)}$$

Let us look at Eqs. (VI.36) rather carefully since some important physical phenomena are inferred by them. The rotational resonance condition $|J_1 - J_2| = 1$ is fulfilled by all four two-molecule eigenfunctions. In Eq. (VI.36a) the individual vibrational quantum numbers remain the same for both system state functions appearing in the matrix element. Equation (VI.36b) presents a different case, however. In this matrix element, the vibrational quantum number for molecule one changes from v to v' under the aegis of U while the vibrational quantum number for molecule two changes from v' to v. Thus, one or more quanta of vibrational energy are exchanged in this process, and resonance in the sense of Chapter V sets in. It seems important to clearly differentiate between this exchange type resonance and the rotational type of Eq. (VI.2). Perhaps this type differentiation is not too satisfactory for $|J_1 - J_2| = 1$ also implies an exchange in that the two molecules may exchange one quantum of rotational energy between themselves. The semantics of the situation should hardly trouble us, however, if we have a clear picture of the physical phenomena involved. In order to determine one of the reasons for neglecting exchange resonance, the portions of Eqs. (VI.36) pertaining to vibration may be written as

$$U_{v1}' = \int \psi_v(1)\mu_1\psi_v(1)d\sigma \int \psi_{v'}(2)\mu_2\psi_{v'}(2)d\sigma = \mu^2 \quad \text{(VI.37a)}$$

$$U_{v1}'' = \int \psi_v(1)\mu_1\psi_{v'}(1)d\sigma \int \psi_v(2)\mu_2\psi_{v'}(2)d\sigma. \quad \text{(VI.37b)}$$

Equation (VI.37a) tells us that $U_i' = U_i$, London's matrix elements. Equation (VI.37b) is proportional to the intensity of the rotation-vibration band involved. For the bands which Lindholm considered, the intensity is very low which means that Eq. (VI.37b) will be small. This in turn will cause Eq. (VI.36b) to be small which allows us to neglect the exchange resonance. Another consideration leads to the weakness of exchange resonance. The dispersion f value enters this resonance broadening as we have seen in Chapter V. In this case, we divide the dispersion f value for a monatomic

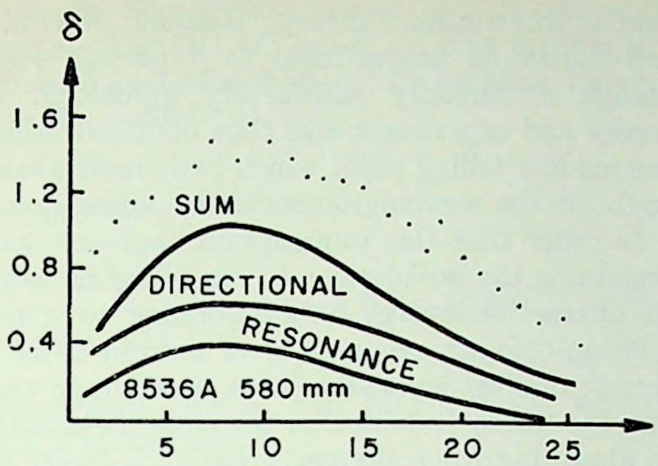


FIG. 17.

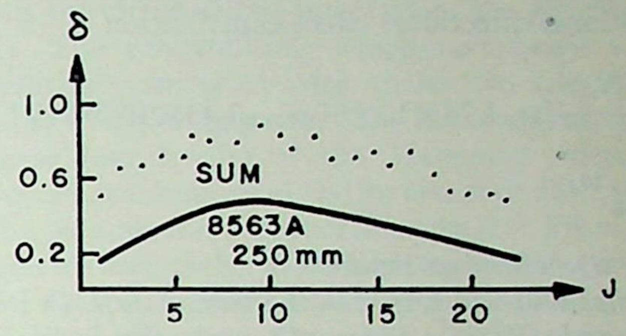


FIG. 20.

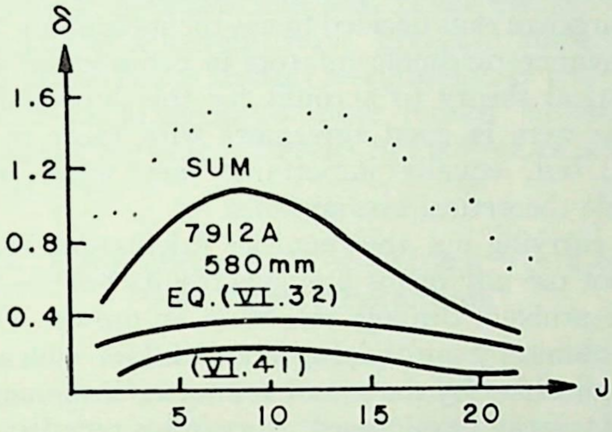


FIG. 18.

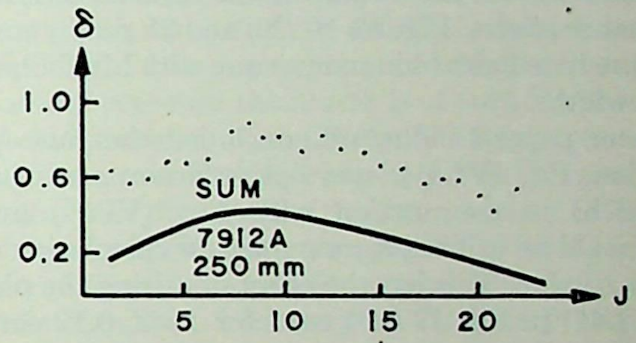


FIG. 21.

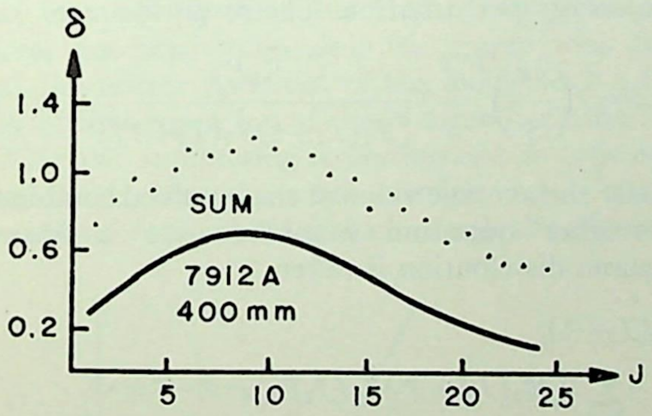


FIG. 19.

FIGS. 17–21. Comparisons of various of Lindholm's theoretical results for the HCN molecule with experimental results. (After Lindholm.[48])

transition among myriads of rotation-vibration levels. The net result of all this is a further decrease in the exchange resonance effect which we now proceed to neglect. This means that we are again only concerned with the twelve matrix elements of Eq. (VI.36a).

A calculation similar to the one leading to Eq. (VI.14) results in

$$R^6E^{(1)} = -\left(\frac{a_1^2}{E_1'}+\frac{a_2^2}{E_2'}+\frac{a_3^2}{E_3'}\right)+\frac{a_4^2}{E^{(1)}-E_4'}, \qquad \text{(VI.38)}$$

where a_i is again given by Eq. (VI.13). This equation is of solution

$$E^{(1)} = \tfrac{1}{2}\left\{E_4 - \frac{\pi^2\mu^4 I}{3h^2R^6} \pm \left[E_4^2 + \frac{2E_4\pi^2\mu^4 I}{3h^2R^6} + \frac{4\mu^4}{6R^6}\right]^{\frac{1}{2}}\right\} \qquad \text{(VI.39)}$$

in which the term in R^{-12} under the radical has been dropped.

The second term within the curly braces in Eq. (VI.39) gives us b_{upper}. b_{lower} is again obtained from Eq. (VI.30), from which

$$\frac{3\pi^2 b}{4} = \frac{3\pi^2}{4}(b_{\text{upper}} - b_{\text{lower}}) = -\frac{7}{8}\frac{\pi^4\mu^4 I}{h^3}. \qquad \text{(VI.40)}$$

It is again necessary to solve Eq. (VI.27b) for ρ.

In his first paper on the subject Lindholm concluded that B contributed little to the optical collision diameter so that he neglected the resonance contribution in this case. This small contribution would indicate that a sharp resonance is required for effect. This sharpness is reduced by the variation of the rotational constant for

the two states since this variation will in turn cause a variation in the rotational level separations for the two vibrational states.

Lindholm then obtained the Case (2) contribution as the additional directional effect contribution

$$\delta=\langle v\rangle\frac{N}{\sum_{J'=0}^{\infty} w_{J'}}[w_J 4.78\times10^{-14}+w_{J-2}3.43\times10^{-14}] \quad (\text{VI.41})$$

with the w_J defined as before.

The total half-width is now the sum of Eqs. (VI.32), (VI.39), and (VI.35). Figure 17 gives the individual directional and resonance effect contributions as well as the sum of the total directional effect plus the resonance effect. Figure 18 gives the individual directional effect contributions as well as the sum of the total directional and resonance effects. Figures 19, 20, and 21 simply give the total predicted widths in comparison with Lindholm's observed widths.

In a later paper Lindholm[49] concluded that his approximation, Eq. (VI.41), was not sufficient and that Eq. (VI.27b) in conjunction with Eq. (VI.39) and (VI.40) should be utilized for a numerical calculation of ρ for each J value. This has the effect of raising the plot of Eq. (VI.41) in Fig. 17 0.06 cm^{-1} for $J=2$, 0.12 cm^{-1} for $J=5$, 0.11 cm^{-1} for $J=10$, 0.05 cm^{-1} for $J=16$, and 0.02 cm^{-1} for $J=20$.

Lindholm also carried out an investigation similar to that of HCN for the HCl molecule. Although his results are rather impressive, the general method of their obtention is precisely the same. Thus, since we have rather thoroughly investigated this method, we shall content ourselves with having mentioned the HCl investigation.

G. An Application of Symmetric Top Dipole Broadening. Ammonia

The development of microwave spectroscopic techniques in recent years has naturally led to a quickened interest on the part of many investigators in pure rotational spectra and the line broadening particular to these spectra. In this and the next several sections we shall consider some of the theories which have been advanced in recent years aimed directly at microwave broadening effects or at microwave and infrared effects.

Bleaney and Penrose[5] have furnished the microwave field with some excellent results on the widths of the ammonia inversion lines. Some of the results of their experiments are contained in Fig. 22. Bleaney[3] used an interruption type approach to this broadening under the assumption that the disturbing ammonia molecule causes a certain perturbation of the absorbing ammonia molecule due to the field of its dipole. A collision is declared as having occurred at a certain separation or interaction energy, and, under an application of the simple interruption theory, Bleaney determines the half-widths as proportional to 3 $[K^2/(J^2+J)]^{\frac{1}{2}}$. Although a cursorily satisfactory agreement between theory and experiment was thus obtained, Margenau[60] has made a telling point which casts more than a little doubt on the meaningfulness of this agreement.

In order that this interruption approach succeed in explaining the broadening, it was found necessary that an interaction energy (corresponding to a minimum collision-defining separation) of around twice the inversion doublet level separation be used. In suggesting this, the theory implies that the molecule could go right on absorbing when the perturbation is larger than the unperturbed level separation, even resulting in a negative frequency. This appears a bit difficult to accept.

Margenau thus decided to use the interaction between two symmetric dipole rotators in conjunction with his statistical theory to account for this broadening. His results were in good agreement with those of experiment, but, equally important, they were based on tenable theoretical assumptions.

In carrying out the requisite calculation Margenau did not use any of his previous results but worked the entire problem through to obtain an answer which we shall obtain by multiplying Eq. (VI.23)—which should be applicable—by the square root of π. We do not repeat the Margenau development since it is a repetition of the obtention of the statistical theory and the root-mean-square energy perturbation. The result is

$$\delta=2\left(\frac{2\pi}{3}\right)^{\frac{1}{2}}\frac{\mu^2}{h}N\frac{K_1K_2}{[J_1(J_1+1)J_2(J_2+1)]^{\frac{1}{2}}}. \quad (\text{VI.42})$$

Finally the average value of the involved combination of perturber quantum numbers over a Maxwell-Boltzmann distribution is taken

$$\langle K_2J_2(J_2+1)\rangle = \frac{\sum_{J_2K_2}[K_2(J_2^2+J_2)]g(J_2K_2)e^{-E(J_2K_2)/kT}}{\sum_{J_2K_2}g(J_2K_2)e^{-E(J_2K_2)/kT}}. \quad (\text{VI.43})$$

In Eq. (VI.43) the $E(J_2K_2)$, are the symmetric rotator energies. The $g(J_2K_2)$ are the statistical weights

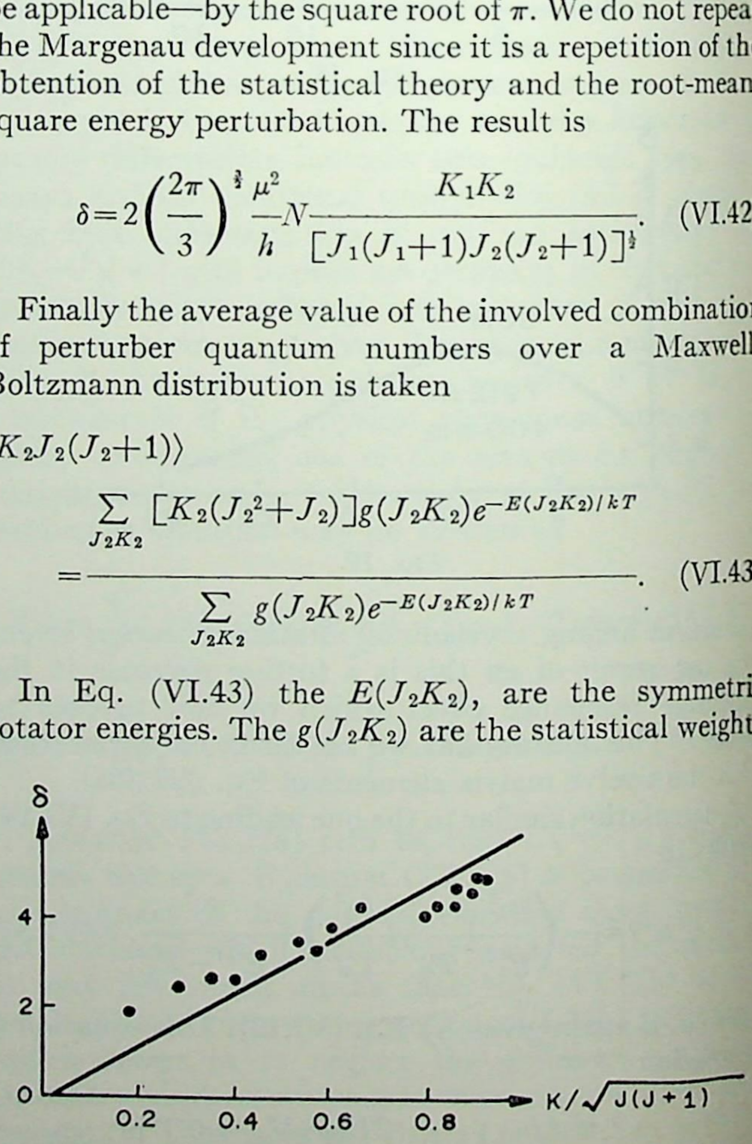


FIG. 22. The line width of the NH_3 inversion line as a function of rotational quantum number. (After Margenau.[60])

of the levels. Smith[78] has carried out the calculation indicated by Eq. (VI.43) to obtain 0.54 for $T=20°C$. For ammonia $\mu=1.44\times10^{-18}$ esu. On this basis Margenau obtained the result

$$\frac{\delta}{p}=1.13\times10^{-3}\times\frac{K_1}{[J_1(J_1+1)]^{\frac{1}{2}}}\text{ cm}^{-1}\text{ (mm Hg)}^{-1}\text{ @ }20°\text{C}, \quad \text{(VI.44)}$$

and the straight line in Fig. 22 yields the results of applying this equation.

This then is Margenau's low pressure (and the theory has not here developed for any but low pressures) result for ammonia inversion lines, and its success for this case is manifestly apparent from Fig. 22. Some of the modifications which higher pressures invoke will become apparent in Margenau's more refined treatment of interactions between potential hill molecules which we consider next.

H. Interactions between Linear Vibrators with Mirror Potentials

The shift to the red with increase of pressure of the NH_3 inversion line, which had recently been observed,[4,78] provoked the Margenau investigation[61] of the phenomenon which we shall now consider. First let us note that in applying the term "linear" to NH_3 (or other mirror potential molecules) we simply wish to imply that the mirror potential of the molecule is a function only of the separation of the N atom from the H plane.

Now two interacting molecules of this type are to be considered. Margenau has carried out a variational perturbation calculation which details the effects of such an interaction. The treatment need not be given, but the results, which are rather informative, can be gleaned by a consideration of Fig. 23. On this figure the two frequencies, into which our inversion frequency has split by virtue of the interaction, are given by Δ_I and Δ_{II}. The probabilities—which correspond to the intensities—associated with these two frequencies also appear on the figure and are indicated as w_I and w_{II}. These four results of the Margenau calculation are plotted as functions of the inversion doublet separation. The parameter Λ is proportional to R^{-3}. From this figure then we may gather that the binary interaction between two such mirror potential molecules splits the inversion line into two components, the most intense of which is shifted by the greatest amount.

The interaction between three molecules with mirror potentials was next considered by Margenau using a variational perturbation technique, and we have indicated the results in Fig. 24. The inversion line has been split into four components of which we have indicated only three since the fourth is of such low intensity as to be not reasonable of display. We might remark that the dominant intensity (frequency-wise) is more strongly so than was the case for the two molecule interaction and becomes dominant more rapidly with Λ. Further, the frequency shifts much more rapidly with Λ than does the dominant frequency in the binary encounter. The trend with higher orders of interaction appears to be apparent then and, we might add, in the proper direction. Thus, this study would indicate that as the pressure increases we may no longer accept the binary approximation as a reasonable one. Let us now consider the low pressure case further.

Under the assumption that the resonance frequency could be approximated by the mean frequency, as given by $h\langle\nu\rangle=w_I\Delta_I+w_{II}\Delta_{II}$ and that Margenau obtained for the shifted frequency

$$\langle\nu\rangle\doteq\nu_0(1-\Lambda^2). \quad \text{(VI.45)}$$

Since

$$\Lambda=\frac{2\mu^2}{(E_1-E_2)R^3}\doteq\frac{2\mu^2N}{E_1-E_2}\doteq0.6p$$

where p is the ammonia pressure in atmospheres, $E_1-E_2\doteq0.8$ cm^{-1} $\mu=1.44\times10^{-18}$ and the temperature is taken as 0°C, Eq. (VI.45) becomes

$$\langle\nu\rangle\doteq\nu_0(1-0.36p^2). \quad \text{(VI.46)}$$

This can only be expected to hold at low pressures since (a) it has been obtained under the assumption of a binary collision or a two particle interaction and (2) the factor Λ has been assumed as small. The "dominant" frequency may also be written approximately for low pressures from the binary theory as

$$\nu_{II}=\nu_0(1-0.6p) \quad \text{(VI.47)}$$

and this may be compared with the experimental results.

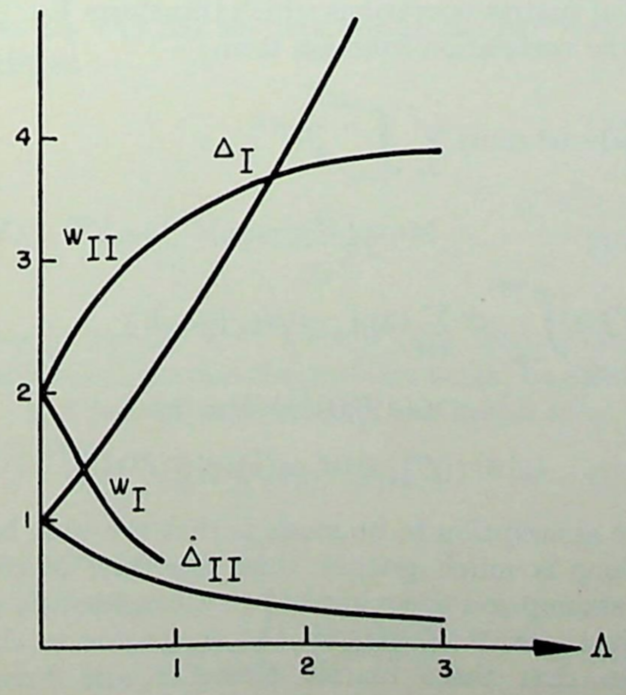


FIG. 23. The splitting and shifts of the NH_3 inversion line due to the binary interaction. Intensities are indicated. (After Margenau.[61])

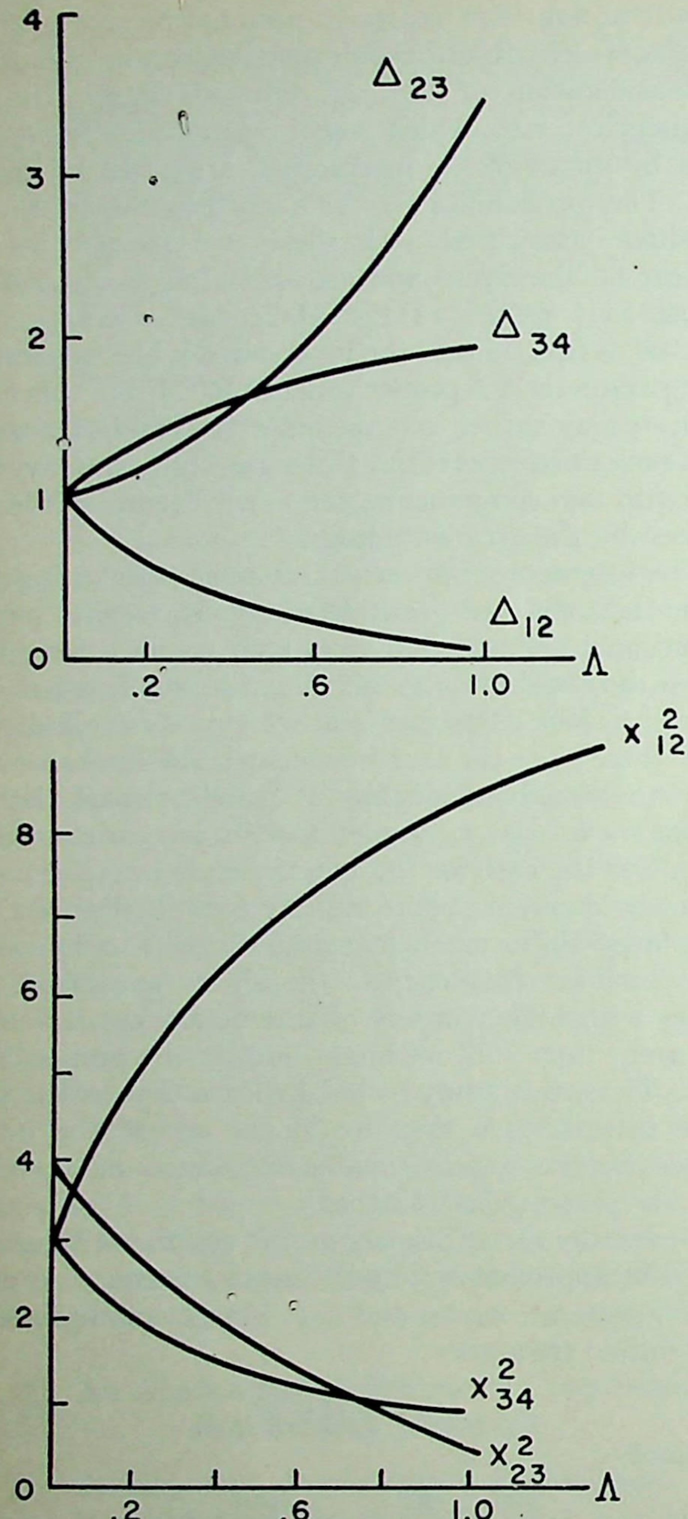


FIG. 24. The splittings of the NH_3 inversion line due to the three particle interaction. (After Margenau.[61])

I. Anderson's Line Broadening Theory

We have made some mention of the adiabatic collision approximation in our considerations of Chapters II and IV, and some additional comments on this subject appear to be necessary here. Let us suppose we are simply considering nondegenerate states which are distorted after some fashion by the collision interaction between broadener and emitter. Now in the optical region of the spectrum the separation of the energy levels of the emitter is so great that it is highly unlikely that this collision interaction will so distort the levels—energy-wise—as to effectively put the atom into another state. As we progress out into the infrared and thence to the microwave the situation changes, however, since the levels between which the contemplated transitions take place move closer and closer together. It follows quite readily that a distortion of these levels may be sufficient to bring about a change of state of the emitter. When the collision causes such a change then we have, of course, experienced a nonadiabatic collision. As a result of all this we can see that, although the adiabatic collision assumption may be a reasonable one in the visible region, it becomes less and less so as we progress into the microwave region. This theory of Anderson's[1] to which we now propose to devote some little consideration, treats nonadiabatic collisions, since it was written for application to the long wavelength transitions of the microwave and infrared regions.

We begin by making the assumption of a classical path. For all save electrons this is a reasonable approximation. We then arrive at the equation for the intensity:

$$I(\omega)=\text{const}\,\omega^4 Tr\left\|\Upsilon_0\int_{-\infty}^{+\infty} dt e^{i\omega t}\mu_z(t)\times\int_{-\infty}^{+\infty} dt' e^{-i\omega t'}\mu_z(t')\right\| \quad \text{(VI.48)}$$

which has a somewhat familiar appearance. In essence a quantum mechanical solution of this equation for the line shape, line width, and line shift under certain rather standard interruption assumptions constitute the residue of the theory.

First the matter of obtaining Eq. (VI.48) must be considered, and we shall simply accept the replacement of the classical Fourier expansion for the dipole moment by an operator and the quantum averaging of the result over a density matrix.

Letting $\|\mu(t)\|=\|U^{-1}\mu_0 U\|$, we then perform a few familiar matrix operations which transform Eq. (VI.48) into the correlation function form:

$$I(\omega)=\omega^4\,\text{const}\sum_{de}\int_{-\infty}^{+\infty} dt'' \times\exp[i(\omega-\omega_{de})t'']\varphi_{de}(t'') \quad \text{(VI.49a)}$$

$$\varphi_{de}(t'')=\int_{-\infty}^{+\infty} dt\sum_{abc}\exp[-i(\omega_{bc}+\omega_{de})t] \times(a|\Upsilon(t)|b)(b|\mu_0|c) \cdot(c|\mathfrak{F}^{-1}(t'')|d)(d|\mu_0|e)(e|\mathfrak{F}(t'')|a). \quad \text{(VI.49b)}$$

The assumption to be made is that the time between collisions is much greater than the time of collision. This assumption when applied to a consideration of the matrix elements of $\|U_0^{-1}H_1U_0\|$ leads one to the conclusion that these matrix elements and hence Eq. (VI.62b) disappear unless

$$\omega_{bc}+\omega_{de}=0. \quad \text{(VI.50)}$$

We specify the density matrix by a typical matrix element as

$$(n|\Upsilon_B|m)\delta_{nm}g_n\exp(-E_n/kT)/\sum_n g_n\exp(-E_n/kT). \quad \text{(VI.51)}$$

We now suppose that different types (this means collisions of different optical collision diameter and direction) of collisions are designated by different values of σ. The probability that a collision will occur during the time interval dt'' and lie in the type range $d\sigma$ is given

$$p(d\sigma \text{ in } dt'')=Nvd\sigma dt''. \quad \text{(VI.52)}$$

Next $d\varphi(t'')$ is averaged over these collisions as

$$d\varphi(t'')=dt''Nv\int d\sigma[\varphi(t''+dt'')-\varphi(t'')] \quad \text{(VI.53)}$$

and $\varphi(t''+dt'')$ and $\varphi(t'')$ may be obtained by utilizing an isotropic collision assumption.

The solution of this equation is

$$G(t'')=e^{-Nv\sigma t''} \quad \text{(VI.54a)}$$

where

$$\sigma=\sigma_r+i\sigma_i. \quad \text{(VI.54b)}$$

In order to obtain the intensity distribution in the broadened spectral line we substitute Eqs. (VI.54a) in Eqs. (VI.49):

$$I(\omega)=\sum_{J_iJ_f}\frac{(J_i|\Upsilon|J_i)}{g_{J_i}}\sum_{M_iM_f}|(J_iM_i|\mu_z|J_fM_f)|^2 \times\frac{Nv\sigma_r}{(\omega-\omega_{fi}+Nv\sigma_i)^2+(Nv\sigma_r)^2}. \quad \text{(VI.55)}$$

From Eq. (VI.55) we may obtain the line half-width and shift as

$$\delta=\frac{Nv}{\pi}\sigma_r \quad \text{(VI.56a)}$$

$$D=\frac{Nv}{2\pi}\sigma_i. \quad \text{(VI.56b)}$$

It would appear that were σ in a form which was amenable to calculation the problem would be essentially solved, and indeed this remains the major step to the solution. Now

$$\sigma=\sum_{J_2}(J_2|\Upsilon|J_2)\sigma_{J_2} \quad \text{(VI.57a)}$$

and we may write σ_{J_2} as

$$\sigma_{J_2}=\int_0^\infty 2\pi\rho d\rho S(\rho), \quad \text{(VI.57b)}$$

where $S(\rho)$ is a trace which is being averaged over all possible optical collision diameters ρ.

In evaluating Eqs. (VI.57) the assumption of a minimum distance of approach—what we might refer to as a cut-off distance—is introduced. Finally, one obtains

$$S=S_1+S_2\cdots \quad \text{(VI.58a)}$$

$$S_1(\rho)=i\left[\sum_{M_iM_2}\frac{(J_iM_iJ_2M_2|P|J_iM_iJ_2M_2)}{(2J_i+1)(2J_2+1)}-\sum_{M_fM_2}\times\frac{(J_fM_fJ_2M_2|P|J_fM_fJ_2M_2)}{(2J_f+1)(2J_2+1)}\right] \quad \text{(VI.58b)}$$

$$S_2(\rho)_{\text{outer}}=\tfrac{1}{2}\left[\sum_{M_iM_2}\frac{(J_iM_iJ_2M_2|P^2|J_iM_iJ_2M_2)}{(2J_i+1)(2J_2+1)}+\sum_{M_fM_2}\times\frac{(J_fM_fJ_2M_2|P^2|J_fM_fJ_2M_2)}{(2J_f+1)(2J_2+1)}\right] \quad \text{(VI.58c)}$$

$$S_2(\rho)_{\text{middle}}=\sum_{\substack{M_iM_i'M_f\\M_f'M_2M_2'\\M}}\sum_{J_2'}\times\frac{(J_f1M_fM|J_f1J_2M_2)(J_f1M_f'M|J_f1J_iM_i')}{(2J_i+1)(2J_2+1)}\times(J_fM_fJ_2M_2|P|J_fM_f'J_2'M_2')\cdot(J_iM_i'J_2'M_2'|P|J_iM_iJ_2M_2), \quad \text{(VI.58d)}$$

where

$$\|P\|=\frac{1}{\hbar}\int_{-\infty}^{+\infty}\|U_0\|^{-1}\|H_c\|\|U_0\|dt \quad \text{(VI.58e)}$$

so that

$$S_2(\rho)=S_2(\rho)_{\text{outer}}+S_2(\rho)_{\text{middle}}. \quad \text{(VI.59)}$$

Let us notice that $S_1(\rho)$ contributes only toward line shift while the $S_2(\rho)$ contributes only to line breadth, not that these facts should prove startling.

As has been the case with the varied and sundry theories which we have considered, the theory breaks down for small values of ρ. This transpires in the following manner. H_c depends—for the dipole-dipole interaction which is the only one which Anderson considered§§§—on r^{-3} in first order and r^{-6} in second order. This means that for sufficiently small r, P will become as large as you like. The expansion of S in terms of P, Eq. (VI.58a), is only valid, however, for small P. Thus does the theory, through S, break down for small ρ. Anderson treats this in a manner which we may now detail.

To begin with we assume that for very small values of ρ the collisions are so strong that they terminate the radiation by causing the molecule to proceed to some different nondegenerate state or when the molecule remains in the same state they result in an arbitrary phase shift that averages to zero. In either case then

§§§ Tsao and Curnutte[84] have also considered dipole-quadrupole and quadrupole-quadrupole interaction within the framework of the Anderson theory.

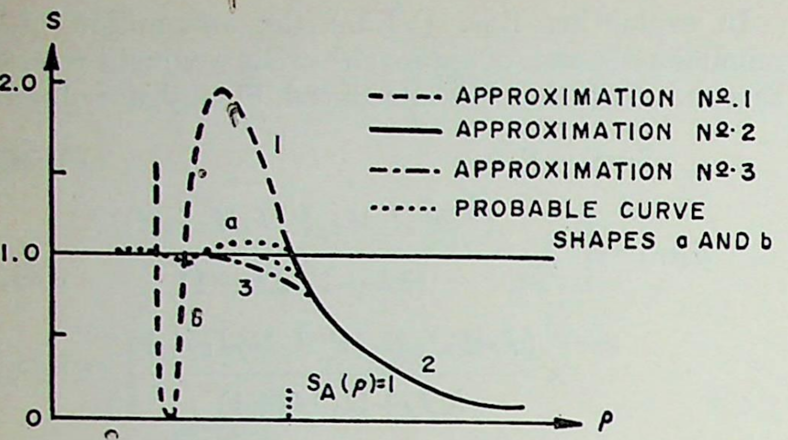


FIG. 25. Anderson's approximation to $S(\rho)$. The probable actual shapes of the curve are also given. (After Anderson.[1])

$S(\rho)$ contains only the first term, unity so that while $S_2(\rho)$ is given by Eq. (VI.76) values of ρ greater than ρ_1 where $S_2(\rho_1)=1$, it is simply given by one for lesser values of ρ.

Anderson also tried two other approximations for $S(\rho)$, namely

$$S_{\#1}(\rho)=1-\cos(-2S_2(\rho))^{\frac{1}{2}}$$

and

$$S_{\#3}(\rho)=1-\exp(-2S_2(\rho)).$$

These three possibilities for S_2 are illustrated in Fig. 25 and we might note here that Anderson found the best agreement with experiment to arise from $S_{\#2}$.

Several applications of this theory have been made, in general with a good measure of success. In some of these application was made of the dipole-dipole interaction which Anderson considered while in others the dipole-quadrupole and quadrupole-quadrupole interactions, which Tsao and Curnutte have studied, was applied. Our approximations here have mainly concerned themselves with (1) the classical path, (2) collision duration very short compared to intercollision time, (3) binary collisions. Approximation (2) is perhaps the least justifiable of the three except at quite low pressures. From our earlier considerations, it is apparent that as the pressures increase, this short collision time approximation becomes a very poor one. Quite recently Tamita[83] has produced a theory which includes the nonadiabatic considerations of Anderson and to them adds a consideration of finite collision time. In general he finds that the Anderson theory is applicable to the line center or low pressure which is what one would expect from our studies of Chapter IV. This latter theory is also primarily concerned with the region of the microwave.

BIBLIOGRAPHY

Periodicals

1. P. W. Anderson, Phys. Rev. **76**, 647 (1949).
2. P. W. Anderson and J. O. Talman, *Conference on the Broadening of Spectral Lines* (University of Pittsburgh, September, 1955).
3. B. Bleaney, Physica **12**, 595 (1946).
4. B. Bleaney and J. H. N. Laubser, Nature **161**, 522 (1948).
5. B. Bleaney and R. P. Penrose, Proc. Roy. Soc. (London) **A189**, 358 (1947).
6. S. Bloom and H. Margenau, Phys. Rev. **90**, 791 (1953).
7. K. G. Carroll, Phys. Rev. **53**, 310 (1938).
8. W. H. J. Childs, Proc. Roy. Soc. (London) **153**, 555 (1936).
9. E. Condon, Phys. Rev. **28**, 1182 (1926).
10. S. D. Cornell, Phys. Rev. **51**, 739 (1937).
11. S. D. Cornell and W. W. Watson, Phys. Rev. **50**, 279 (1936).
12. P. Debye, Physik. Z. **20**, 160 (1919).
13. P. Debye, Physik. Z. **21**, 178 (1920).
14. D. M. Dennison and S. B. Ingram, Phys. Rev. **36**, 1451 (1930).
15. H. Ebert, Ann. Physik **36**, 466 (1889).
16. P. S. Epstein, Phys. Rev. **28**, 695 (1926).
17. H. M. Foley, Phys. Rev. **69**, 616 (1946).
18. J. Franck, Trans. Faraday Soc. **21**, 536 (1926).
19. W. Furssow and A. Wlassow, Physik. Z. Sowjetunion **10**, 378 (1936).
20. W. Furssow and A. Wlassow, J. Phys. (U.S.S.R.) **1**, 335 (1939).
21. R. Gans, Ann. Physik **66**, 396 (1921).
22. W. Grasse, Z. Physik **89**, 261 (1934).
23. C. Gregory, Phys. Rev. **61**, 465 (1942).
24. K. Hedfield and R. Mecke, Z. Physik **64**, 151 (1930).
25. G. Herzberg and J. W. T. Spinks, Z. Physik **91**, 386 (1934).
26. G. Herzberg and J. W. T. Spinks, Proc. Roy. Soc. (London) **A147**, 434 (1934).
27. Herzberg, Spinks, and Watson, Phys. Rev. **50**, 1186 (1936).
28. J. Holtsmark, Ann. Physik **58**, 577 (1919).
29. J. Holtsmark, Physik. Z. **20**, 162 (1919).
30. J. Holtsmark, Physik. Z. **25**, 73 (1924).
31. J. Holtsmark, Z. Physik **34**, 722 (1925).
32. W. V. Houston, Phys. Rev. **54**, 884 (1938).
33. A. Jablonski, Z. Physik **70**, 723 (1931).
34. A. Jablonski, Acta Phys. Polon. **6**, 371 (1937).
35. A. Jablonski, Acta Phys. Polon. **7**, 196 (1938).
36. A. Jablonski, Acta Phys. Polon. **8**, 71 (1939).
37. A. Jablonski, Physica **7**, 541 (1940).
38. A. Jablonski, Phys. Rev. **68**, 78 (1945).
39. H. Kallmann and F. London, Z. physik. Chem. **B2**, 207 (1929).
40. W. H. Keesom, Physik. Z. **22**, 129 (1921).
40a. Kivel, Bloom, and Margenau, Phys. Rev. **98**, 495 (1955).
41. A. Kleman and E. Lindholm, Arkiv Mat. Astron. Fysik **A32**, No. 17 (1946).
42. H. Kuhn, Phil. Mag. **18**, 987 (1934).
43. H. Kuhn, Proc. Roy. Soc. (London) **A158**, 212 (1937).
44. H. W. Kussman, Z. Physik **48**, 831 (1928).
45. W. Lasareff, Z. Physik **64**, 598 (1930).
46. W. Lenz, Z. Physik **25**, 299 (1924).
47. W. Lenz, Z. Physik **80**, 423 (1933).
48. E. Lindholm, Z. Physik **109**, 223 (1938).
49. E. Lindholm, Z. Physik **113**, 596 (1939).
50. E. Lindholm, Arkiv Mat. Astron. Fysik **28A**, No. 3 (1942).
51. E. Lindholm, Arkiv Mat. Astron. Fysik **32A**, No. 17 (1946).
52. W. Lochte-Holtgreven and E. Eastwood, Z. Physik **79**, 450 (1932).
53. H. A. Lorentz, Proc. Roy. Acad. Sci. (Amsterdam) **8**, 591 (1906).
54. C. J. Malmsten, Kgl. Svenska Vetenskapsakad. Handl. **62**, 65 (1841).
55. H. Margenau, Phys. Rev. **40**, 387 (1932).
56. H. Margenau, Phys. Rev. **43**, 129 (1933).
57. H. Margenau, Phys. Rev. **44**, 931 (1933).
58. H. Margenau, Phys. Rev. **48**, 755 (1935).
59. H. Margenau, Phys. Rev. **49**, 596 (1936).
60. H. Margenau, Phys. Rev. **76**, 121 (1949).
61. H. Margenau, Phys. Rev. **76**, 1423 (1949).
62. H. Margenau, Phys. Rev. **82**, 156 (1951).
63. H. Margenau and D. T. Warren, Phys. Rev. **51**, 748 (1937).

64. H. Margenau and W. W. Watson, Phys. Rev. **51**, 48 (1937).
65. A. McKellar and C. A. Bradley, Phys. Rev. **46**, 664 (1934).
66. L. Mensing, Z. Physik **34**, 611 (1925).
66a. R. E. Meyerott and H. Margenau, Phys. Rev. **99**, 1851 (1955).
67. A. A. Michelson, Astrophys. J. **2**, 251 (1895).
68. A. A. Michelson, Phil. Mag. **31**, 338 (1891); **34**, 280 (1892).
69. R. Minkowski, Z. Physik **55**, 16 (1929).
70. R. Minkowski, Z. Physik **93**, 731 (1935).
71. O. Oldenberg, Z. Physik **47**, 984 (1928).
72. O. Oldenberg, Z. Physik **51**, 605 (1928).
73. Lord Rayleigh, Phil. Mag. **27**, 298 (1889).
74. Lord Rayleigh, Phil. Mag. **26**, 274 (1915).
75. P. Schulz and R. Rompe, Z. Physik **108**, 654 (1938).
76. P. Schulz and R. Rompe, Z. Physik **110**, 223 (1938).
77. S-Y. Chen, Phys. Rev. **58**, 884 (1940).
78. D. F. Smith, Phys. Rev. **74**, 506 (1948).
79. L. Spitzer, Jr., Phys. Rev. **55**, 699 (1939).
80. L. Spitzer, Jr., Phys. Rev. **56**, 39 (1939).
81. T. Takamine, Astrophys. J. **50**, 1 (1919).
82. T. Takamine and N. Kokubu, Mem. Coll. Sci. Kyoto Imp. Univ. **3**, 275 (1919).
83. K. Tamita, *Conference on the Broadening of Spectral Lines* (University of Pittsburgh, September, 1955).
84. C. J. Tsao and B. Curnutte, Sci. Rept. IA-8. Contr. No. AF19(122)65 (1954).
85. H. Vedder and R. Mecke, Z. Physik **86**, 137 (1933).
86. S. Verweij, dissertation, Amsterdam, 1936.
86a. S. Verweij, Publ. Astron. Inst. Univ. Amsterdam, No. 5 (1936).
87. J. H. Van Vleck and H. Margenau, Phys. Rev. **76**, 1211 (1949).
88. J. H. Van Vleck and V. F. Weisskopf, Revs. Modern Phys. **17**, 227 (1945).
89. K. Watanabe, Phys. Rev. **59**, 151 (1941).
90. W. W. Watson, J. Phys. Chem. **41**, 61 (1937).
91. W. W. Watson and G. F. Hull, Jr., Phys. Rev. **49**, 592 (1936).
92. V. F. Weisskopf, Z. Physik **75**, 287 (1932).
93. V. F. Weisskopf, Z. Physik **77**, 398 (1932).

Reviews

94. H. Margenau, Revs. Modern Phys. **11**, 1 (1939).
95. H. Margenau and W. W. Watson, Revs. Modern Phys. **8**, 22 (1936).
96. P. Schulz, Physik. Z. **39**, 412 (1938).
97. A. Unsold, Vierteljahrsschr. Astron. Ges. **78**, 213 (1943).
98. V. F. Weisskopf, Physik Z. **34**, 1 (1933).

Books

99. E. Lindholm, *Über die Verbreiterung und Verschiebung von Spektrallinien* (Almquist and Wiksells, Uppsala, 1942).
100. J. Scheiner, *A Treatise on Astronomical Spectroscopy* (Ginn and Company, Boston and London, 1894).
101. J. Stark, *Spstralanalyse Chemische Atome* (S. Hirzel, Leipzig, 1914).

Electromagnetic Structure of Nucleons*

D. R. YENNIE, M. M. LÉVY,† AND D. G. RAVENHALL, *Stanford University, Stanford, California*

The theoretical implications of various experiments relating to the electromagnetic structure of nucleons are examined in the light of current field theory. It is concluded either that the nucleon core is about three times as large as would be expected from intuitive considerations of meson theory, or that there is some inconsistency in the present field theory.

1. INTRODUCTION

INFORMATION about the internal structure of individual nucleons is contained in the results of a variety of experiments performed in recent years.[1] Those experiments in which the interaction with the nucleon is electromagnetic (or is thought to be so) are susceptible of a considerably more precise and unambiguous interpretation than those involving meson interactions. Of the former type (which alone will concern us here) the best known is that involving neutron scattering by atoms, which, when analyzed, gives information about the electron-neutron interaction (to be specific, its volume integral). The Lamb shift and the hyperfine splitting can also give such information, although much less precisely. The fact that the anomalous magnetic moments of the nucleons are equal and opposite has important implications for a meson model of the nucleons, of course. Recent experiments on the scattering of high-energy electrons by hydrogen and deuterium now give considerably more detailed and complete information about the proton and, to some extent, about the neutron. The aim of this paper is to examine the extent to which the results of the separate experiments can be combined into a consistent picture of nucleons as charge-current distributions. We make no claim for the originality of most of the theoretical ideas presented here; they have all appeared in various forms in the literature, and we have brought them together for the purpose of discussing these experiments. In Sec. 2 we present the phenomenology of the electromagnetic interaction of electrons and nucleons. This is interpreted in Sec. 3 in terms of a simple meson model of nucleon structure in which it is assumed that the "physical" nucleon is made up of "bare" nucleons and pions interacting in a charge-symmetrical manner. In Sec. 4 other, more speculative, ways of interpreting these results are suggested, and the implication of these ideas, and their effects on the interpretation of other experiments, are commented on. The Appendix, due to one of us (D. R. Y.), gives a general treatment of the charge-current density of particles of general spin.

* This work was supported in part by the U. S. Air Force through the Air Force Office of Scientific Research, Air Research and Development Command.

† The work was completed while this author was at the University of Paris.

[1] Some of the topics we shall discuss have been reviewed in Bethe and de Hoffmann's book *Mesons and Fields Vol. II* (Row, Peterson and Company, Evanston, 1955), pp. 289–299. In many cases we shall refer the reader to this excellent book for bibliographies.

2. PHENOMENOLOGY OF ELECTRON-NUCLEON INTERACTION

The scattering of an electron from a nucleon caused by their electromagnetic interaction[2] is represented by the Feynman diagram of Fig. 1. We write the matrix element for the process as

$$-4\pi i j_\mu{}^{p,n}(P',P)(1/q^2)j_\mu{}^e(k',k), \tag{2.1}$$

which includes both the Coulomb interaction and the effect of the exchange of transverse photons.[3] The interpretation of this expression is as follows: the factor $(1/q^2)$ represents the propagation of a virtual photon of four-momentum q_μ between the electron and the nucleon where q_μ is the recoil momentum,

$$q_\mu = P_\mu' - P_\mu = -(k_\mu' - k_\mu). \tag{2.2}$$

In the center-of-momentum frame $(q_0=0)$, q^2 is given by

$$q^2 = (2k_c \sin\tfrac{1}{2}\theta_c)^2, \tag{2.3}$$

where k_c and θ_c are the electron's momentum and scattering angle in this frame. The factor $j_\mu{}^e$ is the electron's charge-current density, which, assuming no internal structure, is given simply by

$$j_\mu{}^e(k',k) = -ie\bar{u}(k')\gamma_\mu u(k), \tag{2.4}$$

where $u,\bar{u}$ are Dirac spinors for the electron. The charge-current density of the nucleon $j_\mu{}^{p,n}(P',P)$ (proton or neutron) includes all of the effects of the internal structure.[4] The purpose of the experiments we are discussing

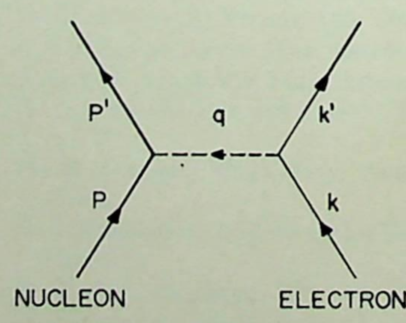


FIG. 1. Feynman diagram of the scattering of an electron by a nucleon caused by the exchange of a virtual photon.

[2] Scattering due to nonelectromagnetic interactions is discussed later in this Section.

[3] R. P. Feynman, Phys. Rev. 74, 939 (1948). Our notation differs from Feynman's somewhat, in that e, the electric charge, is given in unrationalized units, and the four-vector product $a\cdot b$ means $\mathbf{a}\cdot\mathbf{b}-a_0b_0$. Usually we put $\hbar=c=1$.

[4] Technically this is the vertex operator evaluated between free nucleon states.

is to determine information about this charge-current distribution beyond what is already well known from static experiments (the total charge and magnetic moment).

The problem has been simplified by Foldy,[5] and later, in more generality, by Salzman.[5] They show that the nucleon charge-current density must have the form

$$j_\mu^{p,n}(P',P)=ie\bar{v}(P')[\gamma_\mu F_1^{p,n}(q^2) + (\kappa^{p,n}/2M)\sigma_{\mu\nu}q_\nu F_2^{p,n}(q^2)]v(P). \quad (2.5)$$

The assumptions are (i) relativistic covariance, which means that $j_\mu^{p,n}$ transforms as a four-vector; (ii) a differential law of current conservation, which in momentum space is expressed as

$$(P'-P)_\mu j_\mu^{p,n}(P',P)=0; \quad (2.6)$$

and (iii), that the nucleon is a Dirac particle.[6] In Eq. (2.5), the quantities $v,\bar{v}$ are Dirac spinors for the nucleon, $\kappa^{p,n}$ is the anomalous magnetic moment of the nucleon in nuclear magnetons, and the rest of the symbols have their usual meanings. The functions $F_{1,2}^{p,n}(q^2)$ describe the internal structure, and in the static limit ($q^2\rightarrow 0$) take the value unity, except for $F_1^n(0)$, which is zero. As regards the uniqueness of the form of Eq. (2.5), it should be mentioned that there are other covariant expressions satisfying Foldy's assumption, e.g., the convection-current term $(P'+P)_\mu\bar{v}(P')v(P)$. However, such terms always can be expressed in the form (2.5) by the use of the Dirac equation,

$$(\gamma_\mu P_\mu - M)v(P)=0, \quad \bar{v}(P')(\gamma_\mu P_\mu' - M)=0. \quad (2.7)$$

The functions F_1 and F_2 are relativistic generalizations of the form factors characteristic of finite extension occurring in other experiments, for example, in the scattering of electrons from nuclei.[7] There, the form factor is simply the Fourier transform of a radial density function,

$$F(q^2)=\int f(r)\exp(i\mathbf{q}\cdot\mathbf{r})d^3r. \quad (2.8)$$

The function $f(r)$ comes from the product of the initial wave function of the scatterer at rest and the final wave function of the scatterer after it has absorbed the recoil momentum $\mathbf{q}$. For heavy nuclei, where the recoil velocity is negligible compared with c, nonrelativistic wave functions are sufficient, and $f(r)$ is just the static charge or magnetic-moment distribution. For the scattering from a nucleon, however, the situation is qualitatively different, in that for values of $\mathbf{q}$ large enough that the finite nucleon size may be detected, the recoil velocity is comparable with c. It should be emphasized, however, that it is always possible to analyze the experiments in terms of the invariant functions $F_1(q^2)$ and $F_2(q^2)$, and for intuitive convenience to define structure functions $f(r)$ as their Fourier transforms.[8] An accurate calculation of $f(r)$ from some theory would require a correct relativistic description of the internal state of the nucleon. Physically, $f(r)$ would contain the overlap of two wave functions, each Lorentz-contracted, but in different directions. (Actually the structure will be described by a relativistic many-body wave function, so the problem is in fact more complicated than this.) Thus, in relating f to the nucleon wave functions, effects of order v^2/c^2 (or q^2/M^2) are introduced. Consequently there will be a dependence of F on q^2 which is in a sense kinematic in origin, in addition to that coming from the finite extent of the internal wave functions. The essential point is that measurement of structure to within a distance d requires values of $|\mathbf{q}|$ of order $1/d$, and if absorption of this momentum causes relativistic recoil (i.e., if $|\mathbf{q}|>Mc/\hbar$) then intuitive concepts of static charge and current distributions are no longer vaild. Since we expect nuclear structure to extend a distance of order $\hbar/\mu c$, there should be a range of $|\mathbf{q}|$ values ($\mu c/\hbar<|q|<Mc/\hbar$) for which the interpretation in terms of static distributions has some validity. A correct relativistic theory for nucleon structure would, of course, avoid these difficulties by allowing a direct calculation of F as a vertex operator.

To the extent that it is possible to interpret F in terms of static charge-current distributions, it is instructive to make a nonrelativistic reduction of (2.5). This is done in the usual way by expressing the small components of the nucleon spinors in terms of the large components ϕ, which are independent of momentum. In the center-of-momentum frame, and for $|\mathbf{q}|\ll Mc/\hbar$, the components of $j_\mu^{p,n}$ become

$$j_0\simeq e(E/M)\phi_2^*\phi_1[F_1-(q^2/8M^2)(F_1+2\kappa F_2)], \quad (2.9a)$$

$$\mathbf{j}\simeq e[(\mathbf{P}'+\mathbf{P})/2M]\phi_2^*\phi_1F_1 + (e/2M)\phi_2^*(i\boldsymbol{\sigma}\times\mathbf{q})\phi_1(F_1+\kappa F_2). \quad (2.9b)$$

In (2.9b), the first term represents the convection current, and the second the effect of the magnetic moment. In the expression for j_0, in (2.9a), there is, in addition to the expected term F_1, a kinematic term of order q^2 arising from the overlap of the two spinors. The importance of this term in connection with electron-neutron scattering was noted by Foldy.[9]

[5] L. L. Foldy, Phys. Rev. 87, 688 (1952); G. Salzman, Phys. Rev. 99, 973 (1955). An even more general derivation has been given by A. C. Zemach, reference 35. We thank Dr. Zemach for informing us of this work.

[6] This last restriction can be relaxed, as is shown in the Appendix so that Eq. (2.5) holds for any spin one-half particle, e.g. C^{13}. The Dirac spinors are used merely for convenience in representing j_μ in a covariant form.

[7] Hofstadter, Fechter, and McIntyre, Phys. Rev. 92, 978 (1953); L. I. Schiff, Phys. Rev. 92, 988 (1953).

[8] We mean here the three-dimensional Fourier transform, obtained by inverting Eq. (2.8).

[9] L. L. Foldy, Phys. Rev. 88, 693 (1952). It might seem natural to define the spatial distribution of charge as the Fourier transform of the expression in square brackets in j_0. This is not unambiguous, however, as can be seen from Eqs. (A-22) and (A-23).

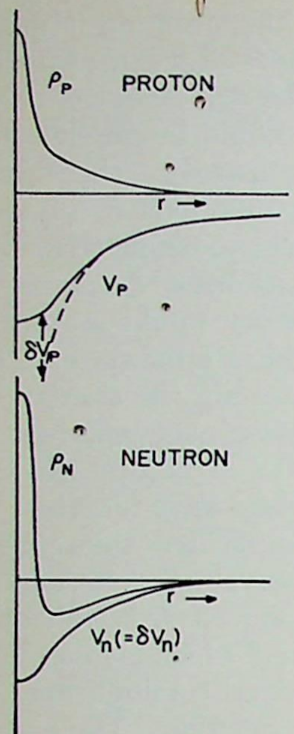


FIG. 2. Intuitive concepts of the neutron and proton charge densities and of the interactions potential between these particles and an electron. The quantities δV represent the departure of these potentials from their values for point nucleons.

For convenience, the functions F can be expanded in powers of q^2. The coefficient of q^2 is simply related to the mean-square radius of the distribution[10]:

$$F(q^2)\simeq 1-q^2\langle r^2\rangle/6+\cdots,\quad \langle r^2\rangle=\int r^2 f(r)d^3r. \qquad (2.10)$$

Because of the relativistic complications discussed previously, there may be contributions to the coefficient of q^2 other than those coming from the finite extension. If this extension is of order $\hbar/\mu c$, they will not completely invalidate intuitive considerations based on static models. With these limitations in mind, we note that the mean-square radii of charge and moment distributions of the nucleons are, from Eqs. (2.9),

$$\begin{aligned}\langle r^2\rangle_{\mathrm{ch}}&=\langle r^2\rangle_1+(3/4M^2)[F_1{}^{p,n}(0)+2\kappa],\\ \langle r^2\rangle_{\mathrm{m}}&=[\langle r^2\rangle_1+\kappa\langle r^2\rangle_2]/(1+\kappa).\end{aligned} \qquad (2.11)$$

In Fig. 2 are shown the usual concepts of the proton and neutron charge densities, and the interaction potentials between those particles and an electron. The quantities δV represent the departure of these potentials from their values for point nucleons, and they can be related to the mean-square radius by Poisson's equation,

$$\begin{aligned}\int\delta V d^3r&=(2\pi e/3)\int r^2\rho(r)d^3r\\ &=(2\pi e^2/3)\langle r^2\rangle_{\mathrm{ch}}.\end{aligned} \qquad (2.12)$$

Electron-Neutron Interaction

Aside from the static limits, the first information about nucleon structure was obtained from experiments on the scattering of neutrons by atoms. Analysis of the results yields the volume integral of the electron-neutron interaction potential, Eq. (2.12). This quantity is conventionally represented by a constant potential, of strength V_0, extending out to a radius $r_0=e^2/mc^2$, the classical electron radius. (This convention is rather confusing since r_0 has nothing to do with this particular interaction.) Hughes *et al.*[11] find for V_0 the value -3860 ± 370 ev, and more recently Melkonian *et al.*[12] report the value -4165 ± 265 ev. Using the mean of these two results, and expressing it in terms of a radius by means of Eqs. (2.11) and (2.12), we find that $\langle r^2\rangle_{\mathrm{ch},n}=-(0.35\times10^{-13}\text{ cm})^2$. This is accounted for completely by the magnetic term of Eq. (2.11), so that

$$\langle r^2\rangle_{1,n}=(0.000\pm0.006)\times10^{-26}\text{ cm}^2. \qquad (2.13)$$

Expressed in terms of potentials, as is customary, the magnetic contribution, first calculated by Foldy,[9] is -4070 ev, leaving for the V_0 associated with $\langle r^2\rangle_{1,n}$ the value 0 ± 200 ev. Had the neutron structure been comparable in extent to that found for the proton, a value of about 0.7×10^{-13} cm would be obtained for $\langle r^2\rangle_{1,n}{}^{\frac{1}{2}}$, or for V_0 about 16 000 ev! The conventional interpretation of this very surprising result is either that the radius associated with the Dirac term is very small, or that the structure, if extended, is almost neutral. In the light of our previous remarks on the meaning of $f(r)$, it is also possible (although perhaps unlikely) that the static charge-current distribution is extended, but that its contribution to $f(r)$ has been canceled fortuitously by relativistic corrections.

Another method for examining neutron structure has been suggested, and is being carried out, by Hofstadter.[13] In the inelastic scattering of electrons from the deuteron at high energies and large angles, binding effects are unimportant, and the neutron and proton scatter independently. Since the electron-proton scattering has been measured separately, the electron-neutron cross section can be deduced. This method can give information about $F_2{}^n$, and possibly also about $F_1{}^n$.

Electron-Proton Scattering

The extensive experiments on the scattering of high-energy electrons by hydrogen of Hofstadter *et al*,[14,15] have given much detailed information about the structure functions of the proton. The analysis uses the formula, first derived by Rosenbluth,[16] for the cross section for the scattering of a relativistic electron by a proton.

[10] Except for $f_{1,n}(r)$, all of the f's have unit volume integral because the corresponding $F(0)$ is unity. Because $F_{1,n}(0)=0$, the quantity $\langle r^2\rangle_{1,n}$ depends on the amount of charge displaced as well as on its radius. For reasonable distributions $\langle r^2\rangle_{1,n}$ will be negative, and the other quantities positive.

[11] Hughes, Harvey, Goldberg, and Stafne, Phys. Rev. **90**, 497 (1953).

[12] Melkonian, Rustad, and Havens, Bull. Am. Phys. Soc. Ser. II, **1**, 62 (1956).

[13] R. Hofstadter, Revs. Modern Phys. **28**, 214 (1956).

[14] R. N. McAllister and R. Hofstadter, Phys. Rev. **102**, 851 (1956).

[15] E. E. Chambers and R. Hofstadter, Phys. Rev. **103**, 1454 (1956).

[16] M. Rosenbluth, Phys. Rev. **79**, 615 (1950).

In the laboratory frame this can be expressed as

$$\sigma(\theta)=\sigma_{NS}(\theta)\{(F_1{}^p)^2+(q^2/4M^2)[2(F_1{}^p+\kappa^p F_2{}^p)^2 \times\tan^2\tfrac{1}{2}\theta+(\kappa^p F_2{}^p)^2]\}, \quad (2.14)$$

where σ_{NS} is the cross section for scattering by a point, spinless particle of mass M:

$$\sigma_{NS}(\theta)=e^4\cos^2\tfrac{1}{2}\theta/\{4k^2[1+2(k/M)\sin^2\tfrac{1}{2}\theta]\sin^4\tfrac{1}{2}\theta\}.$$

The terms in the square bracket in (2.14) arise from the magnetic moment of the proton. The interesting fact that one term involves the total moment while the other contains only the anomalous moment is related to the mixing of F_1 and F_2 in the expressions for charge and current density (2.9).

At the present experimental energies $F_1{}^2$ and $(F_1+\kappa F_2)^2$ are the dominant terms of Eq. (2.14). Since the dependence on angle and energy of the quantity in the curly brackets cannot be expressed in terms of $q(=2k_c \sin\frac{1}{2}\theta_c)$ alone, it is possible to separate the contributions from F_1 and F_2 by performing experiments at various energies and angles. If finite size effects are ignored, so that $F_1=F_2=1$, the first term has the very strong angular dependence and k^{-2} energy dependence characteristic of scattering in a Coulomb field. In contrast, the second term is approximately a constant, independent of energy and angle. Thus the charge scattering and hence the effect of finite charge extension is seen at small angles, while the magnetic moment size is apparent at large angles. The situation is illustrated in Figs. 3. and 4. The "total form factor" $\mathfrak{F}(\theta,k)$ defined by

$$\sigma(\theta)=\sigma_{\text{point}}(\theta)[\mathfrak{F}(\theta,k)]^2, \quad (2.15)$$

where σ_{point} is given by Eq. (2.14) with $F_1=F_2=1$, is plotted as a function of q^2. The three different proton models used for illustration assume either equal charge and moment radii, or else that one of these radii is zero.

McAllister and Hofstadter[14] have measured the elec-

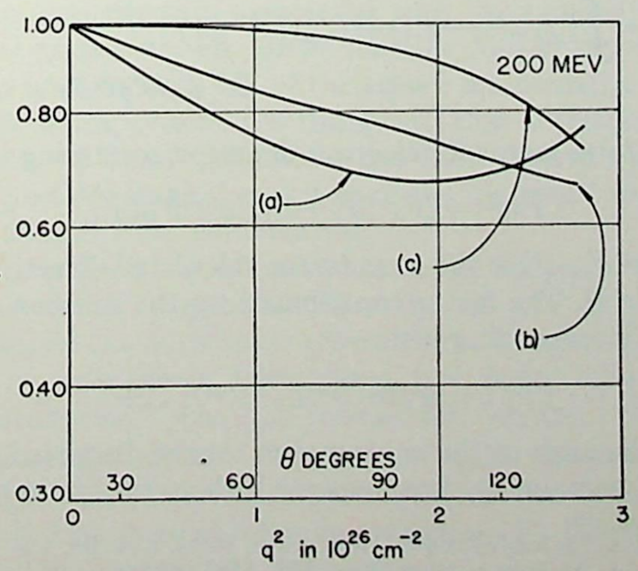


FIG. 3. A plot of the square of the total form factor $\mathfrak{F}$, defined by Eq. (2.15), *vs* q^2, at 200 Mev. Three different proton models are illustrated, corresponding to the following choices for the values of $\langle r^2\rangle_{1,p}{}^{\frac{1}{2}}$ and $\langle r^2\rangle_{2,p}{}^{\frac{1}{2}}$, in 10^{-13} cm: (a) 1.00 and 0.00; (b) 0.70 and 0.70; (c) 0.00 and 0.85.

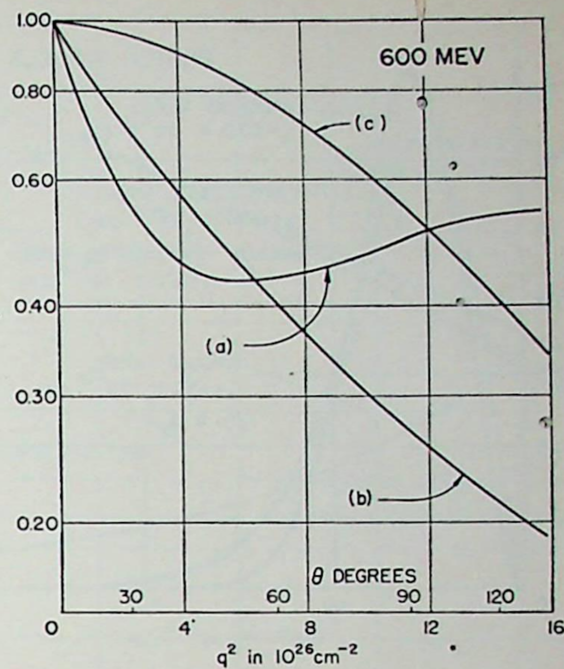


FIG. 4. A plot of $\mathfrak{F}^2$ *vs* q^2 at 600 Mev, for the same cases as in Fig. 3.

tron-proton scattering cross section at 100, 188, and 236 Mev, while more recently experiments at 200, 300, 400, 500, and 550 Mev have been carried out by Chambers and Hofstadter.[15] The detailed analysis of the experiments in terms of F_1 and F_2, and their associated distributions $f_1(r)$ and $f_2(r)$, has been carried out by Hofstadter and his colleagues, and here we shall quote from their results. A feature which makes the fitting with theory a little more flexible than Figs. 3 and 4 might indicate is that the absolute values of the experimental cross sections are not known, although the relative normalization of results for the various energies is known for the latter series of runs. In their analysis, the above authors use the simplifying assumption that the analytic forms of $f_1(r)$ and $f_2(r)$ are identical. The best agreement is found to occur for shapes which are not singular at the center, and which drop off fairly rapidly at large radius. The best fit is then given for equal radii; their actual value depends a little on the choice of shape, but an average value is

$$(\langle r^2\rangle_{1,p})^{\frac{1}{2}}=(\langle r^2\rangle_{2,p})^{\frac{1}{2}}=(0.77\pm0.10)\times10^{-13}\text{ cm}. \quad (2.16)$$

It is possible to have slightly different radii, but neither can be less than about 0.6×10^{-13} cm, or greater than about 1.5×10^{-13} cm. For the case of equal radii the results are illustrated in Fig. 5, where $4\pi r^2 f(r)$ is plotted against r. Chambers[17] has also considered the possibility of different analytic forms for $f_1(r)$ and $f_2(r)$. He finds that it is still not possible to relax the above limits on the radii. The general conclusions are that the proton radii are as given in Eq. (2.16), and that there is no concentration of charge or magnetic moment at the center.

There are possible corrections that might be made to the above analysis, but they are quite unimportant. By

[17] E. E. Chambers, Ph. D. dissertation, Stanford University (1956) (unpublished).

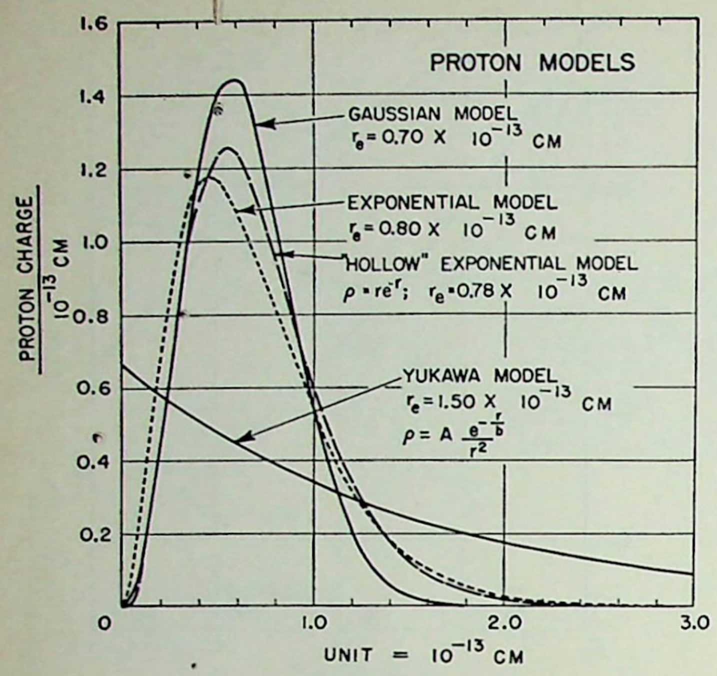


FIG. 5. Graphical representation of proton distribution functions, from the analysis of Chambers and Hofstadter.[15] For the assumption that $f_1(r)=f_2(r)$, these authors find that the Gaussian, exponential and hollow exponential models all give a good fit to the experiments. The Yukawa model, which is suggested by the theory of scalar mesons weakly coupled to nucleons, is given for comparison only; it is in fact inadmissible as a fit to the data. What is plotted *vs* radius r is $4\pi r^2 f(r)$, so that these curves have equal area.

comparing with the results of a partial wave analysis of the scattering process,[18] it is found that the Born approximation used by Hofstadter *et al.* gives cross sections accurate to about 0.25%. The effect on predicted radii is completely negligible. A complication which might influence the interpretation of the form factors F arises from the possibility of virtual excitation of an isobaric state of the proton, corresponding to the resonance in the cross section for photoproduction of mesons. Drell and Ruderman[19] have estimated that the effect of this on the predicted radii is less than about half a percent. The radiative correction of Schwinger[20] has of course been taken into account by Hofstadter *et al.* in their handling of the experimental data.[21] Throughout the analysis it has been assumed that the functions $F(q^2)$ are analytic and "smooth" over the regions of q examined. The electron-proton and electron-neutron experiments are made over completely different ranges of q. It is conceivable, therefore, that the functions F are so peculiar that extrapolation from the one range of q to the other in the simple way that has been assumed is unjustified. A theory that would predict such functions would itself be very peculiar, and we do not consider the possibility further.

Electron-Deuteron Scattering

There exists at present no relativistic theory for the binding of nucleons of finite size to form a nucleus. In the nonrelativistic limit, the solution of the Schrödinger equation describing the interaction of a group of point nucleons is presumably to be interpreted as giving the distribution in space of the centers of mass of the extended nucleons. A difference between this calculated distribution and the measured charge distribution could then be ascribed to finite nucleon size. Use of the nonrelativistic theory unfortunately forces us to ignore the relativistic effects discussed at the beginning of this section.

Experiments on the elastic and inelastic electron-deuteron scattering have been carried out by McIntyre and Hofstadter.[22,23] Following the analysis of Schiff,[24] Jankus[25] has made extensive calculations of both processes, investigating the effect of various assumptions about the neutron-proton potential, and taking into account the effects of magnetic dipole and electric quadrupole moments. Calculations of elastic scattering for particular potentials have been carried out by McIntyre,[23] Bernstein,[26] and Ravenhall.[27] For the detailed comparison of the experiments with these calculations, we quote from the work of McIntyre. He finds that even with the most favorable choice of potential, the experimentally observed charge distribution is significantly more extended than that given by the above calculations. His results are illustrated in Fig. 6.

In the nonrelativistic approximation these calculations give the distribution in space $|\psi_D(r)|^2$ of the centers of mass of the neutron and proton. Neglecting distortions of the nucleon structure due to the binding, the observed deuteron charge density is then

$$\rho(r)=\int[f_{1,p}(|r-r'|)+f_{1,n}(|r-r'|)]\times|\psi_D(r')|^2d^3r'. \quad (2.17)$$

The form factor for electron-deuteron scattering is thus

$$F(q^2)=[F_1{}^p(q^2)+F_1{}^n(q^2)]F_D(q^2), \quad (2.18)$$

where $F_D(q^2)$ is the form factor calculated directly from $|\psi_D(r)|^2$. The factor contributed by the nucleon structure is for small q just

$$[\cdots]\simeq[1-(q^2/6)(\langle r^2\rangle_{1,p}+\langle r^2\rangle_{1,n})+\cdots]. \quad (2.19)$$

On the basis of the naive meson theory discussed in the next section, it was expected that the proton and

[18] Yennie, Ravenhall, and Wilson, Phys. Rev. **95**, 500 (1954), and unpublished calculations.
[19] S. D. Drell and M. A. Ruderman, Phys. Rev. (to be published).
[20] J. Schwinger, Phys. Rev. **76**, 790 (1949).
[21] As in other electron-scattering experiments, the radiative correction is quite large (~20%) but its variation over the angular range used is small (~3%), so that it is not an important correction. The physical situation is different from that considered by Schwinger, in that the proton can emit and absorb photons; but the extra contributions due to this, even that arising from interference with the electron contributions, are negligible.

[22] J. A. McIntyre and R. Hofstadter, Phys. Rev. **98**, 158 (1955).
[23] J. A. McIntyre, Phys. Rev. **103**, 1464 (1956).
[24] L. I. Schiff, Phys. Rev. **92**, 988 (1953).
[25] V. Z. Jankus, Phys. Rev. **102**, 1586 (1956).
[26] J. Bernstein, unpublished calculations. We thank Dr. Bernstein for communication of these results.
[27] Unpublished calculations.

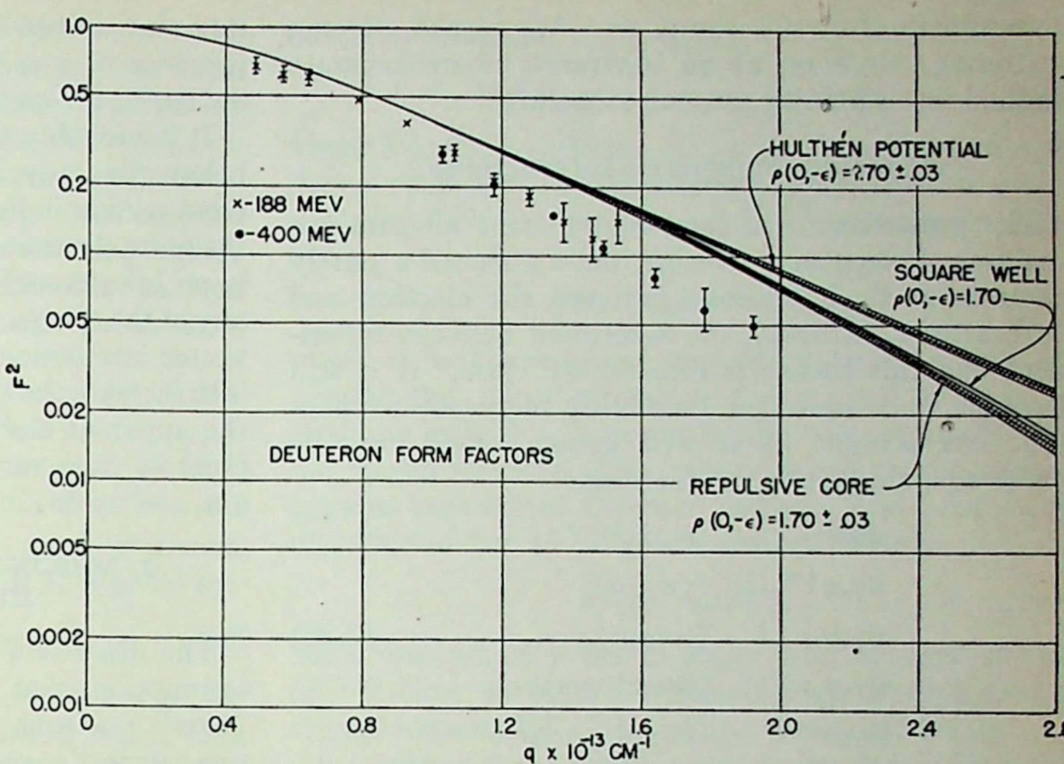


FIG. 6. Experimental and theoretical results of McIntyre on elastic electron-deuteron scattering at 188 and 400 Mev. The theoretical curves include only deuteron S-state contributions; magnetic dipole and electric quadrupole effects are small and would in any case increase F^2. This figure illustrates the fact that the experimental results predict a more extended charge than that obtained assuming point nucleons.

neutron contributions to the q^2 term would cancel, leaving no correction to (2.18) due to the nucleon structure. On the other hand, use of the results of the experiments on the free nucleons means that Eq. (2.19) reduces to the proton form factor. In fact, the analysis of McIntyre strongly favors the latter choice, and can thus be regarded as confirming the difference in size of the neutron and proton. Alternatively, the assumption of nucleon sizes as obtained from the other experiments would give valuable information about the deuteron wave functions.

A possible source of error in the above treatment, associated with the use of the first Born approximation, has been studied by Schiff.[28] In the second Born approximation there is, as well as the usual contribution in which the nucleus stays in its ground state, also a dispersion contribution associated with the possibility of virtual excitation. He shows that the sum of these two effects is of order 1/137 of the elastic scattering, and so is quite unimportant. For the special case of the deuteron, a more exact calculation has been carried out by Volk and Malenka,[29] which confirms his result for the total second Born contribution.

At present, little is known about the accuracy of the nonrelativistic approximation, and of the effect of binding on the nucleon structure. A start to the calculation of relativistic corrections has been made by Blankenbecler,[30] who has treated the deuteron phenomenologically as an elementary vector particle, in analogy with the electron-proton calculation of Rosenbluth.[16] Estimates of mesonic corrections by Bernstein[31] will be discussed in the next section.

[28] L. I. Schiff, Phys. Rev. 98, 756 (1955).
[29] H. S. Volk and B. J. Malenka, Phys. Rev. (to be published).
[30] R. Blankenbecler, unpublished calculation.
[31] J. Bernstein, Phys. Rev. (to be published).

Hydrogen Spectra

There is known to be a discrepancy of 0.6 Mc between the experimental and theoretical values of the Lamb shift for the $2s_{\frac{1}{2}}$ and $2p_{\frac{1}{2}}$ levels of hydrogen and deuterium.[32,33] Part of this can be ascribed to the finite proton size: according to Eqs. (2.11), (2.16), and (2.12), this produces a shift in the $2s_{\frac{1}{2}}$ level of 0.1 Mc,[34] reducing the discrepancy to 0.5 Mc. In deuterium most of the effect of finite size has already been taken into account with a term corresponding to $F_D(q^2)$ of Eq. (2.18).[33] Because of the small neutron size, the finite proton size gives an additional shift of 0.1 Mc in deuterium, the same as in hydrogen.

According to Zemach,[35] the hydrogen hyperfine structure can be used in combination with other experiments to determine a mean electromagnetic radius $\langle r\rangle_{em}$ of the proton: present experimental values lead to the result that $\langle r\rangle_{em}<0.5\times10^{-13}$ cm. This mean radius depends on both the charge and the moment distributions, the relationship among the distributions being

$$\langle r^2\rangle_{em}=\langle r^2\rangle_{ch}+\langle r^2\rangle_{m}.$$

For the shapes predicted by the electron-scattering experiments, which are not peaked at the center, $\langle r\rangle_{em}$ should not be much smaller than $(\langle r^2\rangle_{em})^{\frac{1}{2}}$. The electron-scattering values for the radii predict, however, that $(\langle r^2\rangle_{em})^{\frac{1}{2}}$ is approximately 1.0×10^{-13} cm. There is an apparent disagreement between these two

[32] Triebwasser, Dayhoff, and Lamb, Phys. Rev. 89, 98 (1953); this paper contains references to the earlier work of Lamb and co-workers in this field.
[33] E. E. Salpeter, Phys. Rev. 89, 92 (1953).
[34] The reliability of the perturbation theory has been confirmed by considering exact solutions of the Dirac equation containing a potential of the kind shown in Fig. 2.
[35] A. C. Zemach, Phys. Rev. (to be published). We thank Dr. Zemach for informing us of this work.

determinations of the proton size. We regard electron scattering, however, as an inherently more accurate method for examining nucleon structure.

Nonelectromagnetic Interactions

The preceding discussion, and in fact all previous analyses of electron scattering, have assumed a purely electromagnetic interaction between the electron and the scatterer. Although the agreement between experiment and this theory is remarkably close,[13] it is still desirable to consider the possibility of nonelectromagnetic interactions. As is well known, there are five possible nonderivative interactions,

$$\begin{aligned} &\bar{u}uV^s && \text{(scalar)},\\ &\bar{u}\gamma_\mu uV^v && \text{(vector)},\\ &\bar{u}\gamma_\mu\gamma_\nu uV^t && \text{(tensor)}, \qquad (2.20)\\ &\bar{u}\gamma_5\gamma_\mu uV^{pv} && \text{(pseudovector)},\\ &\bar{u}\gamma_5 uV^{ps} && \text{(pseudoscalar)}, \end{aligned}$$

where the potentials V are of short range. The presence of the vector interaction would not modify the preceding phenomenological analysis, since the finite size effects appear in just this form. For example, the potentials δV of Fig. 2 are $\bar{u}\gamma_4 uV^v$ in this notation. It would alter, however, the interpretation of F_1 and F_2 in terms of nucleon structure; we shall return to this point later. An example of the tensor interaction is given by the electron's anomalous magnetic moment; the Schwinger radiative correction to scattering includes it automatically. For simplicity we will consider first the scalar interaction.

The scalar interaction differs from the interaction with an electrostatic potential by the factor of the Dirac matrix β. For low energy electrons β is effectively unity, and the two interactions will have the same effect. For example, a change of 0.5 Mc in the Lamb shift, mentioned earlier, would require a volume-integral for V^s of 8×10^{-39} Mev cm³. At high energies ($E\gg Mc^2$) it turns out that there is no interference between the two interactions, and the scattering cross section can be written

$$\sigma(\theta)=\sigma_{\rm em}(\theta)+\sigma_s(\theta),$$

$$\sigma_s(\theta)=\left[\int V^s d^3r/4\pi\hbar c\right]^2 q^2[F(q^2)] \times[1+2(\hbar k/Mc)\sin^2\tfrac{1}{2}\theta]^{-2}, \qquad (2.21)$$

where $F(q^2)$ is the form factor for V^s. By comparing (2.21) with (2.14), we see that σ_s has the same form as the term in $\sigma_{\rm em}$ involving the total moment, except for an extra factor of q^2. Thus the effect of σ_s is to reduce the observed magnetic moment size. For example, a scalar interaction with the above volume integral would reduce $\langle r^2\rangle_{\rm m}$ by $(1.7\times10^{-13}\ \text{cm})^2$! If, more plausibly, we were to assume that $\langle r^2\rangle_{\rm m}$ is really $(1.0\times10^{-13}\ \text{cm})^2$, then the fact that the phenomenological analysis has given its value as $(0.8\times10^{-13}\ \text{cm})^2$ could be ascribed to the presence of a scalar interaction with volume-integral 3×10^{-39} Mev cm³.

It is probably true that all of the interactions (2.20) behave similarly, i.e., they would all have effects on the cross section indistinguishable from those coming from the finite electromagnetic sizes. It seems clear that there is no interference between the vector interaction and any of the others. Thus, except for a possible anomalous vector interaction, the effect of the interactions (2.20) is to increase the cross section, and, therefore, to reduce the apparent electromagnetic size. From a theoretical point of view, such interactions would be very unpleasant, and we do not regard the possibility very seriously.

3. MESON-THEORETICAL IDEAS ABOUT NUCLEON STRUCTURE

The discussion of this section will be based on the assumption that "physical" nucleons are made of "bare" nucleons and pions interacting in a charge-symmetrical manner. The results of the phenomeonlogical analysis of the previous section will be examined from this point of view in as general a way as possible. Various calculations of nucleon structure, based on particular meson theories, can then be compared with these conclusions. Some effects associated with the presence of K mesons will also be discussed.

The assumption of charge symmetry implies a relationship between proton and neutron structure. Some of the component states of physical nucleons are represented pictorially in Fig. 7. Charge symmetry requires that, whenever a proton state contains a charged meson, there is a corresponding neutron state with a meson of the opposite sign of the charge. (If the interaction is also *charge-independent*, the amplitudes of the states

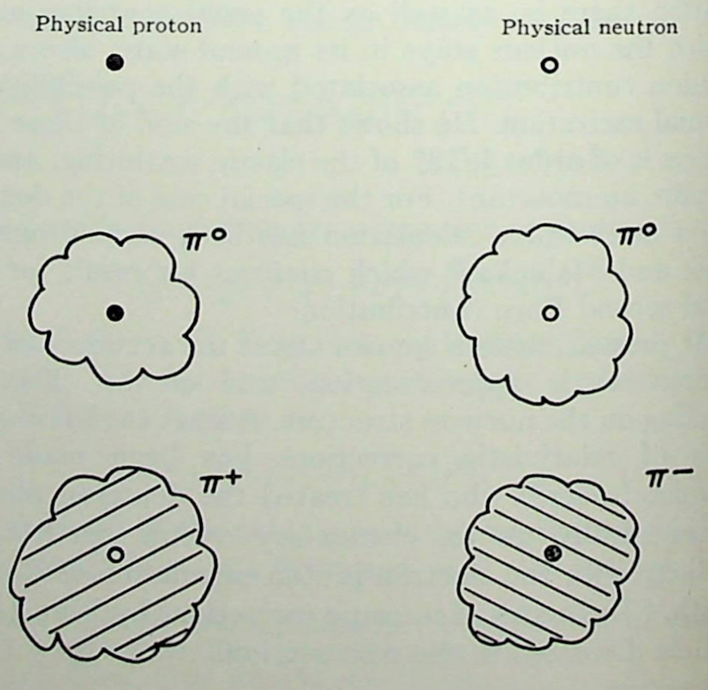


FIG. 7. A pictorial representation of some of the component states of physical nucleons, following the ideas of weak-coupling meson theory. The bare proton is indicated by the black dot, the bare neutron by the open circle.

containing neutral mesons are related to those for the charged mesons.) This is clearly true not only for those component states containing one meson, but for all states. Following Sachs,[36] we can use this result to eliminate the meson part of the nucleon charge density; if we add the proton and neutron charge densities, the meson part cancels completely. The resulting charge distribution, which is a combination of the bare nucleon parts of the proton and neutron charge densities, is called the core distribution;

$$\begin{aligned}\rho_c(r)&=\rho_{c,p}(r)+\rho_{c,n}(r)\\&=\rho_p(r)+\rho_n(r).\end{aligned}\tag{3.1}$$

It includes nucleon pairs as well as single nucleons. From (2.11), the mean square radius of ρ_c is given by

$$\langle r^2\rangle_c=\langle r^2\rangle_{1,p}+\langle r^2\rangle_{1,n}+(3/4M^2)[1+2(\kappa^p+\kappa^n)].\tag{3.2}$$

There may be some doubt about the treatment of the Foldy terms, but fortunately they tend to cancel each other. The experimental values discussed in Sec. 2 then give

$$\langle r^2\rangle_c{}^{\frac{1}{2}}=0.77\times10^{-13}\text{ cm}.\tag{3.3}$$

For comparison, the nucleon compton wavelength is 0.21×10^{-13} cm, and the charge radius of the proton calculated in the above way is 0.84×10^{-10} cm. Thus the core radius is three and one-half times the nucleon Compton wavelength! The results of Chambers and Hofstadter[15] on the shape of the proton charge distribution, which indicate no concentration of charge at small distances, are consistent with this result, and would be inconsistent with a small core radius.

It is very difficult to understand this result. The picture we are considering for nucleon structure would attribute core size to the recoil of the nucleon upon emission of mesons, since the bare particles are assumed to have no intrinsic extension. If the momenta of the emitted mesons are small compared with Mc, by a simple velocity argument the core size will be about one-seventh ($1/7\simeq\mu/M$) of the size of the meson cloud. Under these circumstances the meson cloud extends a distance $\hbar/\mu c$, so that the core should be no bigger than $\hbar/Mc$. If, on the other hand, the virtual mesons are emitted with momenta comparable with Mc, then, because of the relativistic increase in the meson's mass, the core and the meson cloud will be about equal in size, this size being, however, around $\hbar/Mc$. For the states with more than one meson present, the situation is not clear to us. From a semiclassical standpoint, there seem to be a number of possibilities, of which we give two exextremes. As a starting point we assume a recoilless theory in which all of the mesons have the same wave function, centered about the fixed source. A possible approximation to the problem with recoil is to neglect its effect on the state function, and to calculate the core radius by assuming the relation

$$MR+\mu(r_1+r_2+\cdots+r_n)=0,\tag{3.4}$$

where R is the coordinate of the bare nucleon, and r is the coordinate of the ith meson, all measured from the position of the center of mass. It is then easy to show that

$$\langle R^2\rangle=\bar{n}(\mu/M)^2\langle r^2\rangle_{\text{meson}},\tag{3.5}$$

where $\bar{n}$ is the average number of mesons present. Clearly this result cannot be correct if $\bar{n}$ is large. It seems to us that a more reasonable assumption is that the meson wave functions are centered on the instantaneous position of the bare nucleon. The result is to replace $\bar{n}$ of Eq. (3.5) by the average value of

$$n/(1+n\mu/M)^2.\tag{3.6}$$

Since this quantity has an upper limit of about $M/4\mu$, $\langle R^2\rangle$ is close to $(\hbar/Mc)^2$. Although these semiclassical arguments are difficult to justify, the second seems to us the more valid. A correct quantum-mechanical version of Eq. (3.4) will contain field operators, which may link states containing different numbers of mesons, so that the final result may differ from what we have given here. A similar situation with regard to the meson cloud has been pointed out by Sachs.[36] On the other hand, many-meson effects on the meson cloud in the Chew-Low theory do not change the one-meson results qualitatively,[37] and the same result may hold with respect to the core.[38]

[36] R. G. Sachs, Phys. Rev. **87**, 1100 (1952).

[37] S. B. Treiman and R. G. Sachs, Phys. Rev. (to be published); S. Fubini, Nuovo cimento **3**, 1425 (1953); H. Suura (private communication). We are informed by Dr. F. Zachariasen that, due to a computational error, the numerical values given in his calculation [Phys. Rev. **102**, 295 (1956)], which are in disagreement with those of the aforementioned authors, are incorrect.

[38] A different opinion on the size of the nucleon core has been put forward by Tamm [I. Tamm, International Congress on Theoretical Physics, University of Washington (1956), reported by N. N. Bogoliubov] and is as follows. In strong-coupling theory the spatial distribution of nucleon-antinucleon pairs will be closely the same as the meson distribution. If a pair annihilates again, it will not contribute to the nucleon core. Those processes in which the antinucleon of the pair annihilates with the original nucleon do contribute, however, and the large cross section for antiproton annihilation in nuclei observed at Berkeley suggests that they can happen even for pairs created relatively far out in the meson distribution. Thus the presence of many bare-nucleon pairs in the physical nucleon can explain a core distribution which is as large as the meson distribution. These qualitative arguments are, we think, open to question on several counts. That the pair distribution follows the meson distribution will only be true for very strong coupling, where there are so many virtual mesons present that an additional amount of many times $2Mc^2$ added to the energy denominator is inappreciable. Yet the work done on the Chew-Low theory would suggest that the coupling is weak. The application of the experimental results relating to physical nucleon pairs to the annihilation of bare nucleons is not obviously justifiable. Also, it is difficult to believe that the pair contribution can be so closely the same as the meson distribution that the electron-neutron interaction cancels from an expected 10 000 volts down to 200 volts. To settle these questions, it is very necessary to make reliable calculations of the core size. As a final note of comfort, it can be pointed out that if Tamm's suggestion is correct, the true theory is a strong-coupling theory involving the emission of many nucleon pairs, and all meson calculations which have neglected pairs are of little value.

From a theoretical point of view, the correct procedure would be to calculate the appropriate vertex operator with a relativistic field theory, taking account of renormalization. The resulting form factor would include all of the relativistic effects discussed in the last section, and which the above physical arguments ignore. Unfortunately, present techniques allow calculation only in the lowest orders of perturbation theory. A summary of the results of such calculations is contained in Sec. 46 of reference 1. Fried[39] has examined the electron-neutron interaction using pseudoscalar coupling in second order perturbation theory. Using $g^2/\hbar c=13.5$, a "reasonable" value, one finds his formula gives 1300 ev for the specific interaction. Combining this work with Rosenbluth's calculations of electron-proton scattering using the same theory and approximation,[16] one obtains for this value of $g^2/\hbar c$ a core radius of 0.38×10^{-13} cm. Thus the theory to this approximation predicts a not unreasonable value for the electron-neutron interaction, but the core radius is too small by a factor two. The reliability of such calculations can be gauged by looking at their predictions for the anomalous nucleon magnetic moments. There the second-order calculation gives for the ratio $-\mu_n/(\mu_p-1)$ about seven, instead of its observed value close to unity. Although the fourth-order calculations improve the situation somewhat, the results are so different from those in second order as to indicate that no valid conclusions can be drawn from such low-order calculations. It is conceivable that if a correct calculation could be done the results would agree with the experiments. We feel, however, that this is unlikely, and that either the physical model used is basically wrong, or the phenomena can be explained in some other way.

Various calculations of neutron and proton form factors have been made using current cut-off meson theories. In these theories the bare nucleon is treated as a fixed source, and the meson-nucleon interaction is modified to include a cut-off function which suppresses the high momenta. Because the mesons, being pseudoscalar, are emitted into p states, their wave function falls off rapidly outside the source, and its size depends sensitively on the source or cut-off distribution. The cut-off function found by Chew[40] to give agreement with other meson experiments (e.g., meson-nucleon scattering meson photoproduction, etc.) leads to a meson cloud of reasonable size ($\sim0.7\times10^{-13}$ cm). In order to obtain agreement with the observed nucleon sizes, Salzman[41] has assumed arbitrarily that the core extends as far as the meson cloud. As our previous arguments suggest, it seems unlikely that present meson theories can lead to such a large core. A very necessary improvement, which has not yet been made to our knowledge, is the calculation of the core size with this theory. As a first step it may be sufficient to treat the nucleon nonrelativistically.

It has been assumed previously that these static theories do not include the Foldy terms. A proof of this, given by Salzman,[41] involves a two-component reduction of Dirac spinors, taken in the limit as M tends to infinity, but with the anomalous nucleon moment held at its observed value. A different treatment of the charge-current density, given in the appendix, considers charge and magnetic moment as independent quantities. From this point of view the static calculation of the charge density should give all of the interaction, including what from the other viewpoint are called the Foldy terms. It seems to us that all nonrelativistic theories are ambiguous in their treatment of these terms.

An interesting suggestion about the effect of heavy mesons on nucleon structure has been made by Sandri.[42] Because of the "strangeness" selection rule, the K-meson cloud surrounding the nucleon will contain only positively-charged and neutral particles. Thus in adding proton and neutron charge densities to make the core distribution, the K-meson contributions do not cancel. As Sandri points out, s-state mesons extend further than p-state mesons of the same mass. However, the relatively large K-meson mass together with the small coupling constant lead to only a small addition to the core size [about $(\hbar/Mc)^2$ to $\langle r^2\rangle_c$].

The conclusions about the nucleon core are confirmed by the electron-deuteron scattering. In fact, from Eqs. (2.18) and (3.1), we see that the elastic scattering is a process which, so far as its dependence on nucleon size is concerned, automatically measures just ρ_c. Of the various corrections to this simple nonrelativistic picture, Bernstein[31] has considered the fact that the binding is due to the exchange of mesons, rather than to an instantaneous potential. He identifies the state of two physical nucleons, with no mesons being exchanged, with the phenomenological wave function. The additional contribution to the scattering comes from the state in which one meson is being exchanged. The scattering from the exchanged mesons vanishes by charge symmetry; the effect comes from scattering by the nucleons in this state. The nucleon distribution in the one-meson state is much more peaked at the center than for the no-meson state, and as a result the cross section is enhanced for large q. Bernstein calculates the effect to be more than 10% at the largest q values shown in Fig. 6. It is not clear to us what the phenomenological Schrödinger wave functions are an approximation to, and even if the identification with the no-meson state is valid, the calculation has neglected the effect of states containing two mesons on the one-meson state, which is presumably as important as the effect Bernstein has examined. There may well be an effect of this kind, however.

[39] B. D. Fried, Phys. Rev. 88, 1142 (1952).
[40] G. Chew, Phys. Rev. 95, 1669 (1954).
[41] G. Salzman, Phys. Rev. 99, 973 (1955); Phys. Rev. (to be published).
[42] G. Sandri, Phys. Rev. 101, 1616 (1956).

4. ALTERNATIVE INTERPRETATIONS OF THE EXPERIMENTS

Since it is not clear that orthodox meson theory will be able to explain the experimental results, we consider in this section possible alternative interpretations.

Charge Symmetry

Present experimental evidence for belief in charge symmetry and charge independence is discussed in Sec. 30 of reference 1. Charge symmetry is a very plausible hypothesis that, according to experimental results, is clearly true to a good approximation, although it is impossible to make a very accurate quantitative check. A possible solution to the dilemma about nucleon structure is that charge symmetry holds for interactions over large distances, but breaks down completely for short distances. This would imply that the fundamental interaction is not charge symmetric; that the large distance effects are charge symmetric is then difficult to understand. The stronger assumption of charge independence contains charge symmetry. It is used in all current theories of phenomena involving nucleon and pions, and plays a fundamental role in theories of the strange particles. It would be a pity to throw out such a beautiful and simplifying hypothesis if any other way can be found out of the difficulty.

Current Conservation

In deriving Foldy's result, Eq. (2.5), it was necessary to assume a differential law of current conservation. Now we might imagine that, since the nucleon has a complicated structure, charge may not be conserved in small regions, but only as a whole. Although such a modification is objectionable because it violates gauge invariance, for the sake of completeness we consider it briefly.

Current conservation is expressed by Eq. (2.6), which in momentum space says that j_μ should be orthogonal to the momentum transfer q_μ. The simplest and most general charge-current density which violates this condition is

$$j_\mu' = -ieq_\mu \bar{v}(P')v(P). \tag{4.1}$$

In the static limit this implies a radial current flow, but gives no contribution to the charge density. If (4.1) is inserted into the matrix element (2.1), the result is zero because of the conservation law obeyed by the electron. Thus, in order to obtain an electron-proton interaction, we must assume that the electron also violates current conservation. Following this through, we arrive at an additional cross section of the same form as $\sigma_s(\theta)$, shown in Eq. (2.21). Thus, as with the anomalous interactions discussed in Sec. 2, violation of charge conservation will not help in interpreting the experiments.

Electromagnetic Interaction

We can easily explain all of the experiments in terms of a modification of quantum electrodynamics at small distances. The high-momentum cutoffs introduced into the theory by Feynman[43] to suppress the ultraviolet divergences have just this effect. Following Feynman, we may introduce this cutoff as a modification in the photon propagator,

$$1/q^2 \to C(q^2)/q^2. \tag{4.2}$$

From Eqs. (2.1) and (2.5), the effect on the phenomenological analysis is seen to be the replacement of any form factor F by CF; hence these experiments cannot be used to separate finite size effects from those of a possible modification in the Coulomb law. The physical reason for this is that the scattering involves not the charge distribution, whose extension is characterized by F, but its electromagnetic potential V, which depends also on the force law; with the modification (4.2), even a point charge could give a potential of the form shown in Fig. 2.

Because the observed form factors CF_1 and CF_2 in the electron-proton scattering are essentially the same, it is possible to ascribe all of the effects to C, although we expect some intrinsic nucleon size. (If the form factors had been unequal, the two effects would be to some extent distinguishable, in that an upper limit could be put on the radius associated with C, and a lower limit on the actual finite extension.) The modification does not affect the results for the neutron charge radius, since the product $C(q^2)F_{1,n}(q)^2$ is still approximately $(1/6)q^2\langle r^2\rangle_{1,n}$ for small q. For the other experiments—electron-deuteron scattering, Lamb shift, and hyperfine splitting—the effects of finite nucleon size can be reinterpreted as being caused partly or entirely by a modification in the Coulomb law. Such an interpretation makes the reconciliation of the various nucleon properties much easier. With the assumption that, for example, the meson cloud extends only to about $\hbar/Mc$, it is quite plausible that the core and meson distributions are similar enough that the electron-neutron interaction can be as low as 200 ev, especially since the effect of K mesons can be important at this small distance. At the same time the meson clouds in the neutron and proton are charge symmetric, so the near equality of κ^p and $-\kappa^n$ is maintained.

It is hard to devise experiments which would clearly distinguish between a modification in quantum electrodynamics and the effect of finite nucleon size. The most obvious possibility is high-energy electron-electron scattering, but the laboratory energy required to obtain a center-of-momentum energy of 100 Mev is 40 Bev! Processes which involve only real photons, such as Compton scattering, are not affected by a change in the photon propagator [since $C(k^2=0)=1$], although

[43] R. P. Feynman, Phys. Rev. 74, 939 (1948).

particle structure will alter the process. Other electrodynamic processes, such as bremsstrahlung and pair production, usually involve such a small momentum transfer that nuclear size itself is unimportant. In estimating the electromagnetic effects in proton-proton scattering, and in the properties of light nuclei, the two alternatives have slightly different effects, but they would be masked by the greater uncertainty in our knowledge of nuclear forces.

A modification in the Coulomb law would alter slightly the results of other experiments. For example, the radii of nuclear charge distributions deduced from mu-mesic atom level structure and high-energy electron-nucleus scattering would be reduced slightly. For mean square radii the effect is given by

$$\langle r^2\rangle_{\text{obs}}=\langle r^2\rangle_{\text{charge}}+\langle r^2\rangle_c.$$

There would also be some alteration in the nuclear surface thickness.[44]

The theoretical implications of such a modification in quantum electrodynamics have been discussed by Feynman,[45] in an article reviewing the present situation in fundamental theoretical physics, and we will recapitulate some of the points discussed. The renormalization view of the theory regards the cutoffs as mathematical devices to eliminate divergences, with no physical consequences. In contrast to this, Feynman's viewpoint is that they are the manifestation of effects not included in the present theory, which we do not know how to describe in a more fundamental way. According to the former view the theory cannot be used to calculate quantities which depend sensitively on the cutoff—the neutron-proton mass difference, for example. From the latter view, such quantities can be used to give information about the cutoff. Feynman's results for the $N-P$ mass-difference correspond to a cutoff of the same order of magnitude as that required for the electron scattering.

Vacuum polarization affects the photon propagator in the manner indicated by Eq. (4.2), but in the opposite direction to that of a finite size. Feynman brings about the finite size modification described by Eq. (4.2) by introducing "heavy photons." Because their potential must at short distances cancel the Coulomb potential, it is unfortunately necessary that their coupling constant be imaginary. As has been discussed by Feynman, this leads to very fundamental difficulties with regard to conservation of probability. In fact it seems to be impossible to obtain the finite-size effects required from a consistent, point-interaction theory. General arguments lead to an expression for the renormalized photon propagator of the form[46]

$$D_F(q^2)=q^{-2}\left\{1-q^2\int_0^\infty \sigma(\kappa^2)d\kappa^2/\kappa^4(q^2+\kappa^2)\right\}^{-1}, \quad (4.3)$$

where $\sigma(\kappa^2)\geqslant 0$. The form factor must then be of the general form

$$C(q^2)=\left\{1-\int_0^\infty \sigma(\kappa^2)d\kappa^2/\kappa^4 + \int_0^\infty \sigma(\kappa^2)d\kappa^2/\kappa^2(q^2+\kappa^2)\right\}^{-1}. \quad (4.4)$$

Even though it may not be possible to expand the integral of Eq. (4.4) in ascending powers of q^2, it is clear that the whole expression is an increasing function of q^2, and so cannot represent a finite-size effect.

Thus, unless it is possible for pion-nucleon theory to explain the large core size, it seems necessary to make a fundamental revision of present electrodynamic theory. Theoretical arguments concerned with the consistancy of the theory have been advanced by many authors[47] for such a revision, and the nucleon-size experiments may be the first experimental manifestation of this need. It may be necessary to describe nucleons by nonlocal fields, or even to alter our usual concepts of space at small distances.‡

Finite Electron Size

As can be seen from Eq. (2.1), the experiments could also be explained in terms of a finite electron size. Most of the remarks made in connection with the modification of the Coulomb law apply here also. In particular, electron-electron scattering at ultra-high energies would distinguish this possibility from the others. The main objection to this explanation is that there is no reason why the Dirac electron theory should break down at this particular wavelength, since it is clearly valid to wavelengths considerably shorter than $\hbar/mc$, its natural length. We do not regard it as a very likely explanation of the experiments.

[44] For the nuclei examined in Hahn, Ravenhall, and Hofstadter [Phys. Rev. 101, 1131 (1956)], the nuclear surface thickness is decreased by about 13%, while the value of c (the point where the charge distribution has dropped to a half of its central value) is increased by about one percent.

[45] R. P. Feynman, Anais acad. brasil. Ciênc. 26, 1 (1954).

[46] Section 25c of reference 1; G. Källén, Helv. Phys. Acta 25, 417 (1952); H. Lehmann, Nuovo cimento 11, 342 (1954); J. S. Schwinger, Lectures at Stanford, 1956. We would like to acknowledge the clarification of these points produced by Professor Schwinger's stimulating lectures.

[47] G. Källén, *Proceedings of the CERN Symposium* (Geneva, Switzerland, 1956), Vol. 2, p. 187; L. Landau and I. Pomeranchuk, Doklady Akad. Nauk. U. S. S. R. 102, 489 (1955) [a review of the work of Landau and colleagues on this subject is given in Pomeranchuk, Sudakov, and Ter-Martirosyan, Phys. Rev. 103, 784 (1956)]; J. S. Schwinger, reference 46.

‡ *Note added in proof.*—A recent determination of the electron's magnetic moment by P. A. Franken and S. Liebes, Jr. [Phys. Rev. 104, 1197 (1956)] gives the result that $(\mu_e/\mu_0)_{\text{exp}}=1+(\alpha/2\pi)+(0.7\pm2.0)(\alpha^2/\pi^2)$. The value of the third term predicted by quantum electrodynamics [R. Karplus and N. Kroll, Phys. Rev. 77, 536 (1950)] is -2.973 (α^2/π^2). The discrepancy may perhaps be another indication of the breakdown of quantum electrodynamics.

5. SUMMARY

An examination has been made of the present experimental situation regarding the electromagnetic structure of nucleons. It is difficult to understand the remarkable difference in charge radius between the neutron and the proton. Relativistic effects are not expected to be too important, and current meson theories which are charge symmetric seem to us unable to explain the difference. It may be that our physical considerations have leaned too heavily on weak-coupling concepts and results, but a calculation which does not make this approximation, and which at the same time does not neglect recoil, has not yet been made. Apart from this, there seem to be two relatively simple explanations: (i) that charge symmetry does not hold for very small distances; or (ii), that quantum electrodynamics fails at high energies—in other words, that the interaction between two charges is not $(1/r)$ at very small distances. The first alternative would destroy the simplicity of present charge-independent field theories. The second would require a fundamental alteration of present field theory.

ACKNOWLEDGMENTS

In collecting the ideas which form the body of this paper we have benefited greatly from discussions with very many people, both at Stanford and elsewhere. We should like to acknowledge particularly the interest of Professor Robert Hofstadter and Dr. J. A. McIntyre, whose experiments stimulated this investigation. We wish to thank also Professor L. I. Schiff and Professor W. E. Lamb, Jr., for several helpful conversations and Professor Sidney D. Drell for a critical reading of the manuscript.

APPENDIX. GROUP-THEORETICAL TREATMENT OF CHARGE-CURRENT DISTRIBUTIONS OF RELATIVISTIC PARTICLES

There are well-known group-theoretical arguments that nonrelativistic particles of definite spin and parity can possess only certain electric and magnetic multipole moments. In this appendix those arguments will be extended to relativistic particles, and it will be shown in particular that for a spin one-half particle the most general expression of the current density is given by Eq. (2.5).

We shall be interested primarily in matrix elements of the current-density operator between states of definite momentum and spin projection

$$J_\mu(\mathbf{P}',m';\mathbf{P}m)\delta(P'-P-q)=\langle\mathbf{P}',m|j_\mu(q)|\mathbf{P},m\rangle. \quad \text{(A-1)}$$

The physical states of the particle, $|\mathbf{P}m\rangle$, $|\mathbf{P}'m'\rangle$, include all of the effects of the interaction which produces its internal structure; here we need use only the information that these states form a basis for a representation of the inhomogeneous Lorentz group corresponding to a definite spin. The operator $j_\mu(q)$ is the Fourier transform of the current-density operator

$$j_\mu(q)=(2\pi)^{-4}\int j_\mu(x)e^{-iq_\lambda x_\lambda}d^4x. \quad \text{(A-2)}$$

The δ function representing energy and momentum conservation arises from the fact that $j_\mu(x)$ may be expressed

$$j_\mu(x)=e^{iP_\lambda x_\lambda}j_\mu(0)e^{-iP_\lambda x_\lambda}, \quad \text{(A-3)}$$

where P_λ is the energy-momentum operator for the system. In momentum space, charge conservation is expressed by

$$q_\mu j_\mu(q)=0. \quad \text{(A-4)}$$

This imposes a condition on the matrix elements of j_μ

$$(P'-P)_\mu J_\mu(\mathbf{P}'m';\mathbf{P}m)=0. \quad \text{(A-5)}$$

Under a homogeneous Lorentz transformation the state vectors $|\mathbf{P}m\rangle$ and the operators $j_\mu(q)$ transform in a definite way (which fortunately we need consider explicitly only in special simple cases). Thus a knowledge of $J_\mu(\mathbf{P}'m';\mathbf{P}m)$ for one pair of momenta $\mathbf{P}$, $\mathbf{P}'$ and all m, m' will determine J_μ for all other $\mathbf{P}m$, $\mathbf{P}'m'$ that can be reached from the first set by a Lorentz transformation. The totality of such pairs is given by

$$(\bar{P}'-\bar{P})^2=(P'-P)^2=q^2. \quad \text{(A-6)}$$

Thus, for a given q^2, the current density is characterized by at most $4(2s+1)^2$ independent constants: in fact, as we shall see, the actual number of such constants is $(2s+1)$. For a given value of q^2 it now seems appropriate to study the properties of the current density for some particularly simple values of $\mathbf{P}$, $\mathbf{P}'$. The simplest choice seems to be

$$\mathbf{P}'=-\mathbf{P}=\tfrac{1}{2}\mathbf{q}. \quad \text{(A-7)}$$

For simplicity we orient the z axis along $\mathbf{q}$; then since q_0 necessarily vanishes for this combination of momenta, Eq. (5) reduces to

$$J_z(\tfrac{1}{2}\mathbf{q},m';-\tfrac{1}{2}\mathbf{q},m)=0. \quad \text{(A-8)}$$

It will also be convenient to take the direction of spin quantization along the z axis. Then under a rotation about the z axis through an angle ϕ, the states and the operators j_μ will transform according to

$$R_z(\phi)|\pm\tfrac{1}{2}\mathbf{q},m\rangle=e^{im\phi}|\pm\tfrac{1}{2}\mathbf{q},m\rangle \quad \text{(A-9)}$$

$$R_z(\phi)j_0(q)R_z^{-1}(\phi)=j_0(q) \quad \text{(A-10a)}$$

$$R_z(\phi)[j_x(q)\pm ij_y(q)]R_z^{-1}(\phi)=e^{\pm i\phi}[j_x(q)\pm ij_y(q)]. \quad \text{(A-10b)}$$

From this we can easily deduce that (displaying only the m dependence)

$$J_0(m',m)=A(m)\delta_{mm'} \quad \text{(A-11a)}$$

$$J_\pm(m',m)=J_x(m',m)\pm iJ_y(m',m)=B_\pm(m)\delta_{m'm\pm1}. \quad \text{(A-11b)}$$

Further information can be obtained by making the following combination of transformations: (i) space inversion, which leaves m, m' unchanged, but reverses the sign of $\mathbf{q}$ and the spatial components of j_μ; (ii) rotation of 180° about the y axis, which reverses the sign of m, m', but restores the original sign of $\mathbf{q}$ and the x component of j_μ. Since this combination of operations must leave the matrix element unchanged (the specific effect of the parity operation on the states of the particle cancels out because the particle has definite parity), we find the relations

$$A(m)=A(-m)=A(|m|) \tag{A-12a}$$

$$B_\pm(m)=-B_\mp(-m). \tag{A-12b}$$

The minus sign arises from the fact that $R_y(\pi)|\frac{1}{2}\mathbf{q},m\rangle=(-1)^{s+m}|-\frac{1}{2}\mathbf{q},-m\rangle$, according to the usual conventions. The reality properties of the operators

$$j_0^*(\mathbf{q})=j_0(-\mathbf{q})\,;\ \mathbf{j}^*(\mathbf{q})=\mathbf{j}(-\mathbf{q}), \tag{A-13}$$

in combination with a space inversion, lead to the further relations

$$A^*(m)=A(m) \tag{A-14}$$

$$B_\pm^*(m)=-B_\mp(m+1). \tag{A-15}$$

Thus the A's are real. The reality properties of the B's are not fixed by the present considerations alone, but it can be shown by time-reversal arguments[48] that $B_+(m)=B_+[(m+1)$ for the usual choice of phase of the angular momentum states. When this is combined with Eqs. (A-12b) and (A-15), it is seen that the B's are also real. No additional information can be obtained by considering rotations about the x and y axes since, in contrast to the situation in the nonrelativistic case, a preferred axis ($\mathbf{q}$) enters the definition of the spin states.

We may now illustrate these results for the few lowest spins.

Spin Zero

The m label may be omitted, and for the special choice of momenta we have simply

$$J_0=A(q^2), \tag{A-16}$$
$$\mathbf{J}=0.$$

These may be expressed in covariant form,

$$J_\mu(P';P)=(1/2M)(P+P')_\mu F(q^2), \tag{A-17}$$

where

$$F(q^2)=[M/(M^2+\tfrac{1}{4}q^2)^{\frac{1}{2}}]A(q^2). \tag{A-18}$$

This is just the form for the current density of a Klein-Gordon particle modified by a form factor $F(q^2)$, which may represent possible internal structure. We emphasize again that it is not necessary to assume the particle to be a Klein-Gordon particle, but only that it is a spin-zero particle.

Spin One-Half

The nonvanishing matrix elements may be written out explicitly

$$J_0(\tfrac{1}{2};\tfrac{1}{2})=J_0(-\tfrac{1}{2};-\tfrac{1}{2})=A(q^2), \tag{A-19}$$

$$J_+(-\tfrac{1}{2};\tfrac{1}{2})=-J_-(\tfrac{1}{2};-\tfrac{1}{2})=B(q^2). \tag{A-20}$$

In order to write these in covariant form, it is convenient to introduce Dirac spinors which transform in the same way as the states $|\mathbf{P}m\rangle$; these are the usual quantities $v_{\mathbf{P}m}$, which for convenience we take to have the relativistic normalization

$$\bar{v}_{\ m}v_{\mathbf{P}m}=1. \tag{A-21}$$

The form of current density given in Eq. (2.5) has the right transformation properties and the arbitrarines necessary to fit the two functions A and B; it is therefore one possible way of writing Eqs. (A-19) and (A-20) covariantly. Any other covariant forms can always be reduced to Eq. (2.5); this analysis shows also that the most general covariant expression for a current which is not conserved is given by Eq. (4.1). The relation between A, B and F_1, F_2 is

$$\begin{aligned} A(q^2)&=e[F_1(q^2)-\kappa(q^2/4M^2)F_2(q^2)],\\ B(q^2)&=2e(|q|/2M)[F_1(q^2)+\kappa F_2(q^2)]. \end{aligned} \tag{A-22}$$

We may also define a "charge" form factor by

$$F_{\mathrm{ch}}(q^2)=[M/(M^2+\tfrac{1}{4}q^2)^{\frac{1}{2}}]\times[F_1(q^2)-\kappa(q^2/4M^2)F_2(q^2)]. \tag{A-23}$$

The factor $M/(M^2+\frac{1}{4}q^2)^{\frac{1}{2}}$ is the reciprocal of the usual E/M factor in the relativistic charge density [see Eqs. (A-17) and (A-18)].

We can now see why there is an ambiguity in calculating the nucleon's charge distribution according to a fixed source meson theory. In the limit $M\to\infty$, there is no distinction between $A(q^2)/e$, $F_1(q^2)$ and $F_{\mathrm{ch}}(q^2)$; however, for finite M we do not know with which quantity the static charge distribution is to be associated. The difference between the first and second possibilities is just the Foldy term, which is not negligible. The difference between the first and third possibilities is associated with the Lorentz contraction of the charge and it is comparatively unimportant for the analysis of the present paper.

Spin One

A new feature arises in the case of spin one in that it is now necessary to specify two constants in order to

[48] This possibility was suggested by Professor S. D. Drell. For a discussion of time reversal, see S. Watanabe, Revs. Modern Phys. 27, 40 (1955). See also F. Coester, Phys. Rev. 89, 619 (1953).

determine the charge density completely. The new constant is clearly associated with the possibility of an electric quadrupole moment, so we write the charge density in the form

$$J_0(m';m)=A_1(q^2)\delta_{m'm}+A_2(q^2)(m^2-\tfrac{2}{3})\delta_{m'm}. \quad \text{(A-24)}$$

The second term has been so chosen that it vanishes upon averaging over m. Only one constant is needed to specify the current density in the special Lorentz frame.

This result has applications to electron scattering from deuterium. R. Blankenbecler[30] has performed such a calculation by expressing the current density in a covariant form using the β-matrix formalism.

Spin Three-Halves and Higher

One new constant is needed to describe the current density of a spin three-halves particle. It is associated with the possibility of a magnetic octupole moment and occurs explicitly because $J_+(\frac{3}{2},\frac{1}{2})$ and $J_+(\frac{1}{2},-\frac{1}{2})$ are not related to each other by group-theoretical considerations. In this way, every increase of the spin by one-half will result in the possible addition of an electric or magnetic multipole moment.

Cross-Section Formula

In practical applications we often have to evaluate quantities of the form

$$\langle J_\mu{}^*J_\nu\rangle_{\text{Av}}=\frac{1}{2s+1}\sum_{mm'} \times J_\mu{}^*(\mathbf{P}m;\mathbf{P}'m')J_\nu(\mathbf{P}'m';\mathbf{P}m). \quad \text{(A-25)}$$

In the special Lorentz frame it is easy to see that

$$\begin{aligned}\langle J_0{}^*J_0\rangle_{\text{Av}}&=\bar{A}^2,\\ \langle J_0{}^*J_i\rangle_{\text{Av}}&=0,\\ \langle J_i{}^*J_j\rangle_{\text{Av}}&=\begin{cases}\bar{B}^2\delta_{ij}, & i=x,y,\\ 0, & i=z,\end{cases}\end{aligned} \quad \text{(A-26)}$$

where

$$\begin{aligned}A^2&=\frac{1}{2s+1}\sum_m |A(m)|^2,\\ B^2&=\frac{1}{2}\frac{1}{2s+1}\sum_m |B_+(m)|^2.\end{aligned} \quad \text{(A-27)}$$

Defining $\bar{P}_\mu=P_\mu+P'_\mu$, Eq. (A-26) may be expressed in the covariant form

$$\langle J_\mu{}^*J_\nu\rangle_{\text{Av}}=\bar{A}^2[\bar{P}_\mu\bar{P}_\nu/(-\bar{P}^2)] +\bar{B}^2[\delta_{\mu\nu}-(q_\mu q_\nu/q^2)-(\bar{P}_\mu\bar{P}_\nu/\bar{P}^2)]. \quad \text{(A-28)}$$

In order to calculate the cross section, a similar sum must be carried out over the electron spins; the result is

$$\langle j_\mu{}^*j_\nu\rangle_{\text{Av}}\langle J_\mu{}^*J_\nu\rangle_{\text{Av}}=(e^2/2m^2)\{(\bar{A}^2+\bar{B}^2) \times[-q^2-(\bar{P}\cdot\bar{k})^2/\bar{P}^2]+2\bar{B}^2q^2\}, \quad \text{(A-29)}$$

where $\bar{k}_\mu=k_\mu+k_\mu'$. From this the cross section may be obtained directly in any coordinate system by inserting the proper expressions for incident flux, density of final states, etc.

Possible Extensions of the Method

The same method may be applied to the matrix elements for pair production with only changes of detail. In this case $q^2<0$ and it is convenient to choose the frame in which $\mathbf{q}=0$. Then the matrix elements of j_0 vanish, while those of j_z do not. Because the range of q^2 is different from that for scattering, group theory seems to impose no relationship between pair production and scattering. However, if these matrix elements are used directly in higher order perturbation calculations, the result will not usually be gauge invariant (in spite of charge conservation!), and it will be necessary to introduce extra terms to maintain gauge invariance. At present the method seems fruitful only in lowest order perturbation theory.

We may also employ these techniques to calculate the form of the matrix element for systems making a transition (for example, a nuclear transition or photoproduction of pions). For any specific case the procedure would be quite clear, so we shall not present any general rules. Suffice it to say that the result is the relativistic generalization of the usual multipole description of radiative transitions.

REVIEWS OF

MODERN PHYSICS

VOLUME 29, NUMBER 2 APRIL, 1957

International Congress on Theoretical Physics

Held at University of Washington, Seattle, Washington, September 17–21, 1956

THIS issue of the *Reviews of Modern Physics* is devoted to the principal papers presented at the International Congress on Theoretical Physics held at the University of Washington in Seattle, Washington, September 17–21, 1956. This Congress was initiated by the Department of Physics, University of Washington, with the encouragement of the National Science Foundation. Approval was also given by the National Academy of Sciences and the International Union of Pure and Applied Physics. It was made possible by an initial grant from the National Science Foundation, transportation grants from the Office of Naval Research and the Office of Scientific Research, Air Force, and additional funds from the University of Washington and the Research Corporation.

Planning of the Congress was the work of an Advisory Committee consisting of J. R. Oppenheimer and A. Pais, Institute for Advanced Study; L. I. Schiff, Stanford University; F. Seitz, University of Illinois; V. F. Weisskopf, Massachusetts Institute of Technology, and E. P. Wigner, Princeton University, and of a Local Committee at the University of Washington whose members were John S. Blair, E. M. Henley, and Boris A. Jacobsohn. J. H. Manley, University of Washington, served as Executive Secretary of the Advisory Committee. William L. Phillips, University of Washington, served as Congress Coordinator. S. G. Gasiorowicz and R. H. Capps, Radiation Laboratory, University of California, J. R. Schrieffer and R. Thomson, University of Illinois, and F. Mohling, University of Washington assisted in the preparation of the papers for publication.

Participants were effectively saved from brain fatigue by a number of social gatherings: dinners on Monday and Friday evenings, cocktail parties on Tuesday and Thursday, a cruise and salmon bake on Puget Sound on Wednesday and an excursion to Mt. Rainier on Saturday. These essential functions were made possible by generous support of: Avco Manufacturing Corporation, Bell Telephone Laboratories, Boeing Airplane Company, E. I. Dupont de Nemours and Company, Hughes Aircraft Company, General Motors Corporation, Glenn L. Martin Company, Lockheed Aircraft Corporation, Ramo-Wooldridge Corporation, Rand Corporation, Sperry Gyroscope Company, and Westinghouse Electric Corporation.

Attendance included 141 invited participants and 263 observers. Twenty nations were represented. Largest foreign group was that from Japan, consisting of 29 physicists, which gave many participants an opportunity to renew acquaintances made during the previous International Congress of Theoretical Physics held in Tokyo in 1953.

The planners sought successfully to strike a happy medium between the small closed conference for the intensive study of a narrowly defined topic and the completely wide open conference in which a very large number of papers ranges over a vast territory. The program was confined to a few main papers each day, allowing ample time for discussion, and providing for special discussion sessions for consideration of details and short presentation of additional reports by some of the participants. As is to be expected some of the liveliest and most interesting parts of the Congress were the discussions which took place all over Seattle all week long. The Local Committee made arrangements for tape recording and transcribing the main talks and the discussion. However, when decisions had to be made about preparing the material for publication, it did not seem practicable to include the discussion in the published record presented herewith, which therefore includes only the major papers of the program.

In the main outline the program was as follows:

Monday, September 17, morning: E. P. Wigner, chairman. Welcome, by John Manley. Relativity, by V. Bargmann, Princeton University. Nuclear Physics, by V. F. Weisskopf, Massachusetts Institute of Technology. Afternoon: Satio Hayakawa, Kyoto University, chairman. Nuclear Shell Model, by J. H. Jensen, University of Heidelberg. Collective Models of Nucleus, by Ben Mottelson, CERN theoretical study division,

Copenhagen. Nuclear Reactions, by A. M. Lane, Atomic Energy Research Establishment, Harwell, England.

Tuesday, September 18, morning: G. E. Uhlenbeck, University of Michigan, chairman. Statistical Mechanics, by L. Van Hove, University of Utrecht. Statistical Mechanics Panel of short talks by G. E. Uhlenbeck, K. Husimi of Osaka University, L. Onsager of Yale University, I. Prigogine of Free University of Brussels, Jan de Boer of University of Amsterdam, and L. Van Hove. Afternoon: T. Muto, Tokyo University, chairman. Solid State, by C. Kittel, University of California. Superfluidity and superconductivity, by R. P. Feynman, California Institute of Technology. Evening: Discussion groups. Statistical mechanics, led by T. H. Berlin of The Johns Hopkins University. Two-nucleon problem, led by M. M. Levy of the University of Paris.

Wednesday, September 19, morning: J. R. Oppenheimer, Institute for Advanced Study, chairman. Panel on quantum electrodynamics and field theory with short reports by J. R. Oppenheimer, W. Thirring, University of Bern; H. P. Lehman of CERN; E. H. Wichmann of the Institute for Advanced Study; J. Schwinger of Harvard University, S. Tomonaga of Tokyo University of Education; N. N. Bogoliubov of the Academy of Sciences, Moscow.

Thursday, September 20, morning: A. Pais, Institute for Advanced Study, chairman. Meson theories, by H. Yukawa of Kyoto University; π Mesodynamics, by F. E. Low of the University of Illinois. Weak interactions by L. Michel of the Institute of Physics, Lille, France. Afternoon: A. Pais, chairman. New particles, by C. N. Yang, Institute for Advanced Study. Evening: Discussion groups, on low temperature solid state led by R. P. Feynman; on Nuclear Reactions led by L. Rosenfeld, University of Manchester, and on π Mesons led by G. F. Chew of the University of Illinois.

Friday, September 21, morning: Discussion groups, on solid-state problems led by A. Abragam, Center of Nuclear Studies of Saclay, France; Nuclear Models, led by J. A. Wheeler of Princeton University, and Field Theory, Strange Particles and High Energy, led by R. Serber, Columbia University. Afternoon: M. S. Vallarta, National Institute of Scientific Investigation, Mexico City, Mexico, chairman. Origins of cosmic rays, by Phillip Morrison of Cornell University. Stellar energy sources, by E. E. Salpeter of Cornell University.

It is appropriate here to express appreciation of all physicists to the sponsors, to the advisory and local committees and to the contributors to the sessions for their parts in bringing to fruition a most successful congress. Everyone who attended was stimulated with a fresh awareness of the unity of physics and of the many exciting and difficult problems with which the science is confronted today.

It has been said that experts are always bragging about how little they know, the bigger the expert the louder he brags. Those who read the papers which follow will indeed be impressed by the fact that they were prepared by experts.

E. U. CONDON

Relativity

V. Bargmann
Princeton University, Princeton, New Jersey

1. SPECIAL RELATIVITY

1(a) Basic Principles

THE theory of relativity cannot be separated from the rest of physics, for every physical theory is supposed to conform to the basic relativistic principles, and any concrete physical problem involves a synthesis of relativity and some specific physical theory.

To mention one historically important example: When Sommerfeld first derived the fine structure of the hydrogen spectrum, his success was widely considered to be primarily a confirmation of relativistic dynamics. Today we know that the relativistic basis of his analysis was much more secure than the quantum theoretical one. After all, ten years were to pass before the electronic spin was discovered, not to mention wave mechanics or Dirac's theory of the electron.

I shall concentrate here on some of the basic principles relevant to all applications, and on some instances where relativity was crucial for the development of new physical theories.

Special relativity originated from electrodynamics. But only the rudiments of electrodynamics, namely, the fundamental laws of light propagation, were actually used for its foundation. This was the strength of Einstein's position compared to that of Lorentz and Poincaré. Both Lorentz and Poincaré put electrodynamics very much into the foreground, on the assumption that in the last analysis all interactions are electrodynamic interactions (with the possible exception of gravitation). The development of relativity, however, has shown that the theory is in no way restricted to electrodynamics and that it is quite independent of our views on the ultimate nature of the interaction between elementary particles.

For the purpose of this talk, I define the postulate of special relativity as the assertion that in the absence of gravitation the laws of physics have the same form in all inertial frames of reference and that any two inertial frames are connected by a linear coordinate transformation (inhomogeneous Lorentz transformation)

$$x^{i\prime}=a^i+\lambda_j{}^i x^j \quad (i,j=1,\cdots,4) \tag{1.1}$$

which leaves the Minkowskian metric

$$ds^2=-(dx^1)^2-(dx^2)^2-(dx^3)^2+(dx^4)^2 \quad (x^4=ct) \tag{1.2}$$

invariant.

To make this statement unambiguous, several remarks must be added.

(1) Einstein's definition of the space-time coordinates—in particular, his definition of simultaneity—is explicitly assumed. Thus the postulate contains the law of light propagation in free space as well as relativistic kinematics.

(2) In a given inertial frame, K, the state of a physical system must be given by a well-defined set of quantities, say, $f_1, \cdots, f_n$ (which may or may not be functions of space-time), and the transformation law of the f_α must be explicitly stated. This means that in a second inertial frame, K', the same state is described by n similar quantities f'_α which are related to the original f_α by n equations

$$f'_\alpha=F_\alpha(f_1,\cdots,f_n), \tag{1.3}$$

where the form of the functions F_α is determined by the Lorentz transformation connecting K and K'. (Dirac's equation shows how intricate the transformation law may prove to be.) I list here indiscriminately all quantities describing the system, such as hydrodynamical and electromagnetic ones, and the equations (1.3) imply in no way a close relationship between the various f_α.

(3) Passive and active Lorentz transformations. The transformations considered so far—relating the descriptions of the same physical situation in two different frames K and K'—may be termed *passive* Lorentz transformations. Since the laws of physics have the same form in K and K', it follows that the f'_α describe also a possible situation in K, different, in general, from that described by the f_α. This leads to an interpretation of (1.3) as an *active* transformation. If, in a fixed frame K, $\{f_1,\cdots,f_n\}$ is an admissible state of the physical system in question, so is $\{f'_1,\cdots,f'_n\}$.

(4) Finally, not all Lorentz transformations need be admitted. The minimal requirement is that a theory be invariant with respect to the "restricted" Lorentz group, i.e., the group of all transformations which involve neither time reversal nor spatial reflexions. [In terms of the coefficients in (1.1) these are defined by: (i) $\lambda_4{}^4>0$, (ii) the determinant of the $\lambda_j{}^i$ is equal to 1.]

As is well known, three other Lorentz transformations may be adjoined to the restricted Lorentz group:

(*A*) $x^{i\prime}=-x^i\ (i\neq 4);\ \ x^{4\prime}=x^4$ (reflexion)

(*B*) $x^{i\prime}=x^i\ (i\neq 4);\ \ x^{4\prime}=-x^4$ (time reversal)

(*C*) $x^{i\prime}=-x^i\ \ (i=1,\cdots,4)$. (total inversion)

Note that C is the product of A and B.

The distinction of the different possibilities is particularly important for the "active" interpretation. Thus invariance with respect to time reversal expresses the reversibility of the phenomena described by a given theory. In the past one has considered at least the group of all transformations that do not invert the time (i.e., the "restricted" group together with A), but the recent observations on "strange" particles have led to the suggestion of dropping[1]—tentatively—the spatial reflexion A.*

1(b) Experimental Verification

Regarding the experimental verification of relativistic kinematics, I refer to Robertson's analysis of the observational basis of the Lorentz transformations.[2] It concerns three basic optical experiments, and it is, in my opinion, a particular virtue of this analysis that it depends as little as possible on any specific physical theory. As in the definition of the Lorentz group only the basic laws of light propagation are involved. The three experiments are (1) the Michelson-Morley experiment, (2) the experiment by Kennedy and Thorndike[3] (whose arrangement is similar to that of the Michelson Morley experiment, with the distinction, however, that the two arms along which the light is propagated are as different in length as feasible), (3) the experiment by Ives and Stilwell on the transverse Doppler effect, which exhibits the time dilation required by the Lorentz transformation. Robertson has shown that under very modest assumptions one can deduce from the results of these experiments—to the extent that such a thing can be deduced at all—the validity of the Lorentz transformation (and, hence, of relativistic kinematics) including terms of order v^2/c^2.

Apart from these optical experiments, application and verification of special relativity have, of course, been mostly connected with elementary particles. This holds in particular for relativistic dynamics as distinguished from relativistic kinematics.

A most impressive confirmation of the kinematic effect of time dilation is provided by a comparison of the decay time of π mesons in flight with that of π mesons at rest.[4]

Today there is no longer any need to stress the experimental verification of relativistic dynamics. Particles of extreme relativistic energies are an everyday occurrence in physical laboratories, and no failure of relativistic dynamics is known. This applies specifically to the most important consequence of relativity, namely, the equivalence of energy and mass.

1(c) Significance of Relativity for Quantum Theory

Let me mention a few instances where relativity has been crucial for the development of quantum theory. I might start with the very notion of a photon, for the relation $E=cp$ between the energy and the momentum of a particle can certainly not be understood on the basis of Newtonian mechanics. (Planck's book on heat radiation contains a proof that light cannot be assumed to consist of particles, because such an assumption would be incompatible with the relation $P=\frac{1}{3}u$ between the radiation pressure P and the density of radiation energy, u. Planck's analysis rests, of course, on the Newtonian formula $E=p^2/2m$.)

The second instance I want to mention is the discovery of de Broglie's relation. It was the requirement of Lorentz invariance that led from Planck's equation $E=h\nu$ to the relation $p=h/\lambda$.

The third is the inauguration of relativistic quantum mechanics by Dirac. Not only did Dirac's equation provide the correct description of the kinematics and the dynamics of the electron, but it also led to the concept of antiparticles and to the prediction of the creation and annihilation of pairs, thus turning the apparent paradox of negative energy states into the most remarkable success of the theory.

Finally, all of quantum field theory belongs here, because relativity has played an increasingly important role in its development, particularly during the last ten years.

In addition to these instances, where relativity has been of vital importance for the discovery or the formulation of new laws of physics, there are some general results which give a deeper insight into the structure of quantum theory. Firstly, the connection between Lorentz invariance and conservation laws. Conservation laws follow whenever the field equations are derived from a Lorentz invariant Lagrangian. (Admittedly, this connection is not specifically quantum theoretical, but is equally valid for a classical field theory.) Secondly, Pauli's analysis of the relation between spin and statistics,[5] which is based on Lorentz

---

[1] See C. N. Yang's report on new particles.

* *Note added in proof.*—In the meantime the nonconservation of parity has been brilliantly and successfully demonstrated. By themselves, however, the experiments do not prove that spatial reflexion must be omitted from the invariance group of special relativity, because—as was pointed out in the previous remarks—the invariance of a theory with respect to a given Lorentz transformation can only be adequately discussed if the transformation laws (1.3) are explicitly stated for *all* physical quantities. According to the highly satisfactory interpretation of the recent experiments by Lee, Oehme, and Yang (to appear in Phys. Rev.) the theory remains invariant if spatial reflexion is coupled with charge conjugation. Thus the change in our theoretical views does not lead to the omission of spatial reflexion from the list of admissible Lorentz transformations, but it rather leads to a radical alteration of those transformation laws (1.3) which correspond to spatial reflexion, an alteration involving the transition from particles to antiparticles. Since heretofore charge conjugation by itself had been considered a symmetry operation of the theory the total number of symmetry operations has been cut in half.

[2] See reference 12 (bibliography).

[3] See H. P. Robertson's remark in the discussion following this talk, p. 173.

[4] See reference 13 (bibliography). Here the factor $(1-\beta^2)^{-\frac{1}{2}}$ is of the order 1.5.

[5] See reference 14 (bibliography).

invariance, and thirdly Wigner's investigation of the representations of the inhomogeneous Lorentz group, which resulted in the classification of all free relativistic systems that obey the basic laws of quantum theory.[6]

Relativity did not start out as a specifically atomic theory, but it has survived the transition from classical to quantum theoretical concepts. This, I believe, is not a coincidence, for those laws of physics which express a basic "invariance" or "symmetry" of physical phenomena seem to be our most fundamental ones. (Even in Newtonian physics Euclidean geometry—and, consequently, also the group of orthogonal transformations—plays a much more important role than the traditional expositions seem to indicate.)

The present difficulties of relativistic quantum theory are familiar, and it is not known how they will be resolved. If we shall have to abandon the Minkowskian description of the space-time continuum the notion of Lorentz invariance will no longer be directly applicable. But there is no doubt that even then the basic relativistic postulate of the equivalence of all inertial frames will remain valid in some—suitably modified—form.

2. GENERAL RELATIVITY

This second part is concerned with the classical theory of general relativity as originally formulated by Einstein. Various attempts to go beyond this theory are surveyed in Parts III and IV.†

I shall analyze the logical and mathematical structure of general relativity in much greater detail than I did in the case of special relativity in Part I. For one thing, special relativity is an every-day working tool of many theoretical physicists, whereas general relativity is not. For another, many of the details are relevant to the discussion in Parts III and IV.

Familiarity with the elements of Riemannian geometry is assumed—to the extent that they are treated in Einstein's "Meaning of Relativity." I also follow Einstein's notation.

2(a) Basic Principles

General relativity starts with Einstein's equivalence principle. It follows from the equality of inertial and gravitational mass that for purely mechanical phenomena a uniformly accelerated frame of reference is equivalent to a frame at rest which carries a uniform gravitational field. ("Acceleration" and "rest" are taken here with respect to an inertial frame of Newtonian mechanics.) Einstein's principle postulates the equivalence of such frames for *all* physical phenomena and thereby makes gravitation an integral part of general relativity. The analysis of even the simplest cases shows that coordinate differences are no longer directly measured by clocks and measuring rods. Thus, one is led to the introduction of general coordinate systems for the local description of the space-time continuum.

General relativity rests, then, on the following assumptions:

I. General covariance. The laws of physics have the same form in all coordinate systems.

II. Local validity of special relativity. The laws of special relativity hold locally in a coordinate system with vanishing gravitational field.

To these must be added

III. The precise form of the field equations.

Let me begin the analysis with assumption II.

First of all it is meant to imply the existence, in the neighborhood of a given point in the space-time continuum, of a frame of reference in which no gravitational field is present. In this Lorentz frame we have a Minkowskian interval

$$ds^2 = -(dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (dx^4)^2 \qquad (2.1)$$

which is measured by "ideal" clocks or measuring rods according as $ds^2 > 0$ or $ds^2 < 0$. In an arbitrary coordinate system this interval takes the form

$$ds^2 = g_{ik}dx^i dx^k \quad (g_{ik} = g_{ki}) \qquad (2.2)$$

the components g_{ik} of the metric tensor being functions of the coordinates.[7] Here is the origin of the Riemannian structure of the space-time continuum, which is fundamental for general relativity.

Secondly, in a Lorentz frame the world line of a test particle is a time-like straight line in the absence of external forces, so that $d^2x^i/ds^2 = 0$. For a general coordinate system this leads to the *principle of the geodesic line*: In a purely gravitational field, the world line of a test particle is given by the equations

$$\frac{d^2x^i}{ds^2} + \Gamma_{jk}{}^i \frac{dx^j}{ds}\frac{dx_k}{ds} = 0 \qquad (2.3)$$

$$g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 1. \qquad (2.3a)$$

Here, $\Gamma_{jk}{}^i$ are the coefficients of the *affine connection* defined by the metric tensor,

$$\Gamma_{jk}{}^i = \frac{1}{2}g^{im}\left(\frac{\partial g_{mj}}{\partial x^k} + \frac{\partial g_{mk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^m}\right) \qquad (2.4)$$

and s/c is the proper time of the particle. In a given coordinate system, the gravitational field (which determines the motion of a particle) is determined by

[6] See reference 15, also 16(a), (b).

† *Note added in proof.*—The July issue of the *Reviews of Modern Physics* will contain a number of papers on various aspects of general relativity, in particular on the quantization of covariant theories, a question which is only briefly discussed in part IV of the present article.

[7] In a very interesting paper C. Møller discusses the question to what extent real physical systems may approximate "ideal" clocks (reference 17, bibliography).

the coefficients Γ and, hence, by the derivatives $\partial g_{ij}/\partial x^k$. The g_{ik} themselves act as the potentials of the gravitational field and thus appear in their dual role describing both the metrical properties of the space-time continuum and the dynamical action of gravity.

In the neighborhood of a given point P_0 of the space-time continuum a coordinate system ("geodesic" coordinate system) can be introduced such that at P_0 (1) the line element has the Minkowskian form (2.1) and (2) all coefficients $\Gamma_{jk}{}^i$ vanish (i.e., the g_{ik} are constant in first approximation). This well-known mathematical fact clearly demonstrates the dependence of the gravitational field on the frame of reference, and it also permits us to express the "local validity" of special relativity in a precise mathematical form. The coordinate system referred to in assumption II is such a geodesic coordinate system, for which the gravitational field vanishes at P_0.

With the help of the principle of general covariance the equations of special relativity can then be "translated" into general relativity in an unambiguous way if we assume the general relativistic equations to contain no higher than first-order derivatives of the g_{ik}, and if no new quantities, in addition to the metric tensor, are introduced to describe the system in question.[8]

Examples are Maxwell's equations, the equations of hydrodynamics; furthermore, the equations of motion for a charged particle in an external electromagnetic field, which are obtained by inserting the relativistic Lorentz force on the right-hand side of (2.3). (One might also mention the world line of a light ray, which is a null geodesic as follows from Maxwell's equation.)

2(b) Field Equations of Gravitation

To establish the field equations we must go beyond special relativity. In the case of a weak static gravitational field, (2.3) go over into the Newtonian equations of motion if one sets in first approximation

$$g_{44} \sim 1 + 2\phi/c^2, \tag{2.5}$$

where ϕ is the classical potential of Newtonian gravitation. The general relativistic equations ought to be obtained by generalizing Poisson's equation

$$\nabla^2\phi = 4\pi\gamma\rho$$

(γ = Newtonian gravitational constant, ρ = mass density of the sources of the gravitational field). In view of the special relativistic principle of the equivalence of mass and energy ρ must be replaced by the total energy density of the sources, or rather by the total (symmetric) energy-momentum tensor T_{ij} of the sources. Hence, we are led to a tensor equation of the form

$$H_{ij} = T_{ij}, \tag{2.6}$$

where the components of the symmetric tensor H_{ij} are universal functions (i.e., the same in every coordinate system) of the g_{ik} and their derivatives. To maintain the analogy with Poisson's equation the following—very modest—assumption is added: H_{ij} contains no higher than second-order derivatives of the g_{ik}, and it is linear in the second derivatives. The mathematical analysis shows that H_{ij} has then necessarily the form

$$H_{ij} = a_1 R_{ij} + a_2 g_{ij} R + a_3 g_{ij},$$

where R_{ij} is the contracted curvature tensor, R the curvature scalar, and the a_i are constants.[9]

In a Lorentz frame the energy-momentum tensor satisfies the conservation law,

$$\partial T^{ij}/\partial x^j = 0, \tag{2.7}$$

which implies for a general coordinate system

$$T^{ij}{}_{;j} = 0. \tag{2.8}$$

Here, and in the sequel, the semicolon indicates covariant differentiation. In view of the assumed field equations (2.6) we must then have $H^{ij}{}_{;j} = 0$, which is generally true if and only if $a_2 = -\frac{1}{2}a_1$, for in every Riemannian space the curvature tensor satisfies the "Bianchi" identities

$$G^{ij}{}_{;j} = 0 \quad (G_{ij} = R_{ij} - \tfrac{1}{2} g_{ij} R). \tag{2.9}$$

Dividing by a_1, we obtain finally

$$G_{ij} + \Lambda g_{ij} = -\kappa T_{ij} \tag{2.10}$$

(with $\Lambda = a_3/a_1$, $\kappa = -1/a_1$). Comparison with Poisson's equation for a weak gravitational field (produced by a mass distribution of low density) shows that

$$\kappa = 8\pi\gamma/c^4$$

if T_{ij} has the dimension of an energy density.

This completes the derivation of the field equations to the extent that they are determined by the assumptions made. The cosmological constant Λ—which has the dimension 1/(length)²—remains undetermined. Einstein introduced it—reluctantly—in 1917 in order to account for a static universe on the basis of general relativity, but later, particularly when the idea of a static universe had to be abandoned, he strongly advocated discarding it. I believe that many physicists share this view today. Although the astronomers will probably have the last word on this question, I may indicate some of the objections: (1) A universal length $\Lambda^{-\frac{1}{2}}$ is introduced whose value is in no way determined by the theory. (2) Even in the absence of any sources ($T_{ij} = 0$) the Minkowskian metric form is no longer a solution of the field equations.—In any event, the length $\Lambda^{-\frac{1}{2}}$ has cosmic dimensions so that Λ may be disregarded in all applications outside of cosmology. Accordingly, I shall omit it throughout most of the following discussion.

---

[8] Only if such a qualification is added does the principle of general covariance have an unambiguous meaning. [See also the discussion in 1(a).]

[9] See Pauli, reference 2 (bibliography) art. 17, p. 598.

In a complete description of a physical system, T_{ij} will depend on other field variables (electromagnetic, hydrodynamic, etc.) which in turn satisfy an additional set of field equations (Maxwell's, Euler's etc.)

2(c) Consequences of the Covariance of the Field Equations

The covariance of the relativistic field equations has striking consequences. I shall confine myself to the gravitational equations, but what I have to say holds—with minor modifications—for any covariant set of equations, in particular for any general relativistic unified field theory.

Continuation properties[10]

Consider a pure gravitational field ($T_{ij}=0$) with vanishing cosmological constant. Then

$$G_{ij}=0.$$

Given the "initial values" g_{ik} and $\partial g_{ik}/\partial x^4$ on the hypersurface S which, in a suitable coordinate system, has the equation $x^4=0$. To what extent do the field equations determine the continuation of the g_{ik} in the x^4 direction? The covariance of the equations implies immediately that the continuation cannot be *uniquely* determined, for if $g_{ik}(x)$ is a solution with the given initial values, so is any field $g'_{ik}(x)$ obtained from $g_{ik}(x)$ by a (continuous) change of the coordinate system outside S.

If, everywhere on S, $g_{44}\neq 0$ (geometrically: if S nowhere touches the local light cone $g_{ik}dx^i dx^k=0$) the covariance of the theory may also be utilized as follows: Assuming the x^4 direction time-like, we may introduce a coordinate system in which

$$g_{44}=g^{44}=1; \quad g_{4\alpha}=g^{4\alpha}=0 \quad (\alpha=1,2,3).$$

Then we are left with ten equations ($G_{ik}=0$) for only six unknown functions $g_{\alpha\beta}(\alpha,\beta=1,2,3)$. One finds that six equations (*viz.*, $G_{\alpha\beta}=0$) determine the continuation in the x^4 direction, i.e., they may be solved for the derivatives $\partial^2 g_{\alpha\beta}/(\partial x^4)^2$, while the remaining four ($G_{4\alpha}=0$, $G_{44}=0$) are conditions to be satisfied on a section $x^4=\text{const}$.

Finally, once the conditions $G_{4\alpha}=0$, $G_{44}=0$ are satisfied for $x^4=0$ and the $g_{\alpha\beta}$ are determined from the six equations $G_{\alpha\beta}=0$, the field equations must insure the validity of $G_{4\alpha}=0$, $G_{44}=0$ also outside the hypersurface S. That this is indeed the case follows from the Bianchi identities $G^{ij}{}_{;j}=0$ [see Eq. (2.9)] which, in the coordinate system we have chosen, imply $\partial\mathfrak{G}_{4\alpha}/\partial x^4=0$, $\partial\mathfrak{G}_{44}/\partial x^4=0$, where $\mathfrak{G}^{ij}=|g|^{\frac{1}{2}}G^{ij}$ (g=determinant of the g_{ik}).

It should be emphasized that these are purely local considerations. Strictly speaking, they hold for a sufficiently small part of S and for sufficiently small values of x^4. Little is known on the very difficult question of how far the g_{ik} may be continued free of singularities.

Identities and Conservation Laws[11]

The covariance of the field equations, by itself, does not insure the existence of the necessary identities. But if the field equations are derived from a covariant variational principle, the identities as well as the conservation laws may be deduced. This fact demonstrates the importance of a variational principle for covariant field equations.

Einstein's gravitational equations (2.10) are obtained from

$$\delta\int \mathfrak{L}d^4x=0; \quad \mathfrak{L}=\mathfrak{R}-\Lambda|g|^{\frac{1}{2}}-\kappa\mathfrak{L}_s. \tag{2.11}$$

Here, g is the determinant of the g_{ik}, $|g|$ its absolute value, $\mathfrak{R}=|g|^{\frac{1}{2}}R$, and $\mathfrak{L}_s$ is the nongravitational part of the Lagrangian, the subscript s referring to the sources of the gravitational field. For pure electromagnetism,

$$\mathfrak{L}_s=\tfrac{1}{2}|g|^{\frac{1}{2}}f_{ik}f^{ik}; \quad f_{ik}=\frac{\partial\phi_i}{\partial x^k}-\frac{\partial\phi_k}{\partial x^i}. \tag{2.11a}$$

The quantities to be varied are (1) the components g_{ik} of the metric tensor and (2) whatever additional field variables are contained in $\mathfrak{L}_s$ (e.g., the four components of the vector potential in the electromagnetic case). Variation of the g_{ik} yields the gravitational equations, variation of the other variables the remaining field equations (for example, Maxwell's equation). As usual, the variations of the field variables (and, if necessary, of some of their derivatives) are supposed to vanish at the boundary of the fixed four-dimensional region over which the integration is extended.

I shall illustrate the procedure for the derivation of identities for the pure gravitational field, i.e., for $\mathfrak{L}=\mathfrak{R}$ (setting $\Lambda=0$), the essential fact being that $\mathfrak{R}$ is a scalar density.[12]

We first define the following transformation of fields. (1) Assume that with every point x of some region in space-time a point y is correlated by the equations

$$y^i=y^i(x) \quad \text{or} \quad y=y(x).$$

(2) This point transformation induces a corresponding transformation of scalars, vectors, etc.: A scalar field

[10] See reference 3 (bibliography), Chap. II, in particular pp. 31–32.

[11] See reference 2 (bibliography), art. 23.

[12] While the essential features of the method are correctly described here the particular form (B), (C) of the conservation laws is obtained if we follow the customary procedure of replacing $\mathfrak{R}$ by a Lagrangian $\mathfrak{L}'$ which contains only first-order derivatives of the g_{ik}. $\mathfrak{L}'$ is formed by subtracting from $\mathfrak{R}$ a suitable divergence: $\mathfrak{L}'=\mathfrak{R}-\partial\mathfrak{K}^i/\partial x^i$. [See reference 2, art. 23 or reference 4, p. 196 (bibliography).] $\mathfrak{L}'$ is no longer a scalar density, but the variation $\delta\mathfrak{L}'$ induced by (2.14) is still a divergence. The comparison of (2.14a) and (2.14b) remains the crucial step, but the variation of $\mathfrak{R}$ must now be expressed by the variation of $\mathfrak{L}'$ and of $\mathfrak{K}^i$.

$\phi(x)$ is transformed into

$$\phi'(x)=\phi(y(x)),$$

a covariant vector field $v_i(x)$ into

$$v_i'(x)=\frac{\partial y^k}{\partial x_i}v_k(y(x)), \qquad (2.12)$$

contravariant vectors and tensors of all ranks and types being transformed accordingly.

$\mathfrak{R}$ is a universal function of the g_{ik} and their derivatives. If the "transformed" values $g'_{ik}(x)$, $\partial g'_{ik}/\partial x^l$ etc., are inserted, one obtains

$$\mathfrak{R}'(x)=\mathfrak{R}(g_{ik}',\partial g'_{ik}/\partial x^l,\cdots)=J\cdot\mathfrak{R}(y(x)) \qquad (2.13)$$

(J is the Jacobian $\partial(y^1,\cdots,y^4)/\partial(x^1,\cdots,x^4)$) because $\mathfrak{R}$ is a *scalar density*. Incidentally, it is precisely at this point that the covariance of the theory is used.

Introducing a family of transformations

$$y^i(x,\epsilon)=x^i+\epsilon\xi^i(x),$$

we find that, for a given choice of the functions ξ^i, the transformed field variables are turned into functions of x and ϵ. Their variations are the first-order terms of their expansions in powers of ϵ, thus [from (2.12)]

$$\delta v_i=\epsilon\frac{\partial}{\partial\epsilon}v_i'(x,\epsilon)|_{\epsilon=0}=\epsilon\left(\frac{\partial\xi^k}{\partial x^i}v_k+\frac{\partial v_i}{\partial x^k}\xi^k\right)$$
$$\delta g_{ij}=\epsilon\left(\frac{\partial\xi^k}{\partial x_i}g_{kj}+\frac{\partial\xi^k}{\partial x^j}g_{ik}+\frac{\partial g_{ij}}{\partial x^k}\xi^k\right). \qquad (2.14)$$

The variation of $\mathfrak{R}$ may now be expressed in two ways. On one hand,

$$\delta\mathfrak{R}=\frac{\partial\mathfrak{R}}{\partial g_{ij}}\delta g_{ij}+\frac{\partial\mathfrak{R}}{\partial g_{ij,k}}\frac{\partial}{\partial x^k}(\delta g_{ij})+\cdots \qquad \left(g_{ij,k}=\frac{\partial g_{ij}}{\partial x^k}\right)$$

or, using the standard device of variational calculus,

$$\delta\mathfrak{R}=-\mathfrak{G}^{ij}\delta g_{ij}+\frac{\partial\mathfrak{B}^i}{\partial x^i} \qquad (2.14a)$$

($\mathfrak{G}^{ij}$ are the left-hand sides of the field equations, $\mathfrak{B}^i$ are expressions linear in the variations δg_{ij} and their derivatives.) On the other hand, since $\mathfrak{R}$ is a scalar density [see (2.13)],

$$\delta\mathfrak{R}=\epsilon\left(\frac{\partial\xi^i}{\partial x^i}\mathfrak{R}+\frac{\partial\mathfrak{R}}{\partial x^i}\xi^i\right)=\epsilon\frac{\partial}{\partial x^i}(\xi^i\mathfrak{R}). \qquad (2.14b)$$

The last two expressions for $\delta\mathfrak{R}$ coincide for every choice of ξ^i if the variations (2.14) are inserted in (2.14a). Equating the coefficients of the ξ^i and of their derivatives of various orders in (2.14a) and (2.14b) one obtains, therefore, a series of relations which hold *identically*, i.e., whether or not the field equations are satisfied.

In the case of $\mathfrak{R}$ these relations may be reduced to the following three sets:

$$\text{(A)}\quad \mathfrak{G}_{i;j}^{j}\equiv\frac{\partial\mathfrak{G}_i^j}{\partial x^j}-\frac{1}{2}\frac{\partial g_{jk}}{\partial x^i}\mathfrak{G}^{jk}\equiv 0. \quad (\mathfrak{G}_i^j=|g|^{\frac{1}{2}}G_i^j)$$

$$\text{(B)}\quad \frac{\partial}{\partial x^j}(\mathfrak{G}_i^j-\mathfrak{U}_i^j)\equiv 0.$$

$$\text{(C)}\quad \begin{cases}\mathfrak{G}_i^j-\mathfrak{U}_i^j\equiv-\dfrac{\partial\mathfrak{A}_i^{jk}}{\partial x^k},\\[2mm] \mathfrak{A}_i^{jk}=-\mathfrak{A}_i^{kj}=|g|^{-\frac{1}{2}}g_{il}\dfrac{\partial}{\partial x^m}(\mathfrak{g}^{jl}\mathfrak{g}^{mk}-\mathfrak{g}^{jm}\mathfrak{g}^{lk}).\end{cases}$$

I turn now to the interpretation of these relations. (A) are the Bianchi identities whose significance for the field equations I have discussed already. The equations (B) constitute *conservation laws* for energy and momentum. (The $\mathfrak{U}_i^j$ are quadratic in the gravitational field).[13] Using the gravitational field equations (with $\Lambda=0$)

$$\mathfrak{G}_i^j=-\kappa\mathfrak{T}_i^j=-\kappa|g|^{\frac{1}{2}}T_i^j$$

we may rewrite (B) in the form

$$\frac{\partial\mathfrak{S}_i^j}{\partial x^j}=0;\quad \mathfrak{S}_i^j=\mathfrak{T}_i^j+\mathfrak{t}_i^j, \qquad (2.15)$$

where $\mathfrak{T}_i^j$ is the energy-momentum density of the gravitation producing field, while $\mathfrak{t}_i^j=-\kappa^{-1}\mathfrak{U}_i^j$ may be considered the gravitational part of the total energy-momentum density. The $\mathfrak{t}_i^j$ form a "pseudo-tensor" density, i.e., they transform like the components of a tensor density only under *linear* coordinate transformations. (At any given point in space-time, the $\mathfrak{t}_i^j$ vanish in a geodesic coordinate system, in which the gravitational field vanishes.)

Regarding the equations (C) we see, first of all, that the conservation equations (B) follow from them, in view of the antisymmetry of $\mathfrak{A}_i^{jk}$ in j and k. Furthermore, they greatly facilitate and clarify the transition to the conservation laws for total energy and momentum. These may be defined for an isolated system whose metric is "asymptotically Minkowskian," so that, in a suitable frame of reference, the g_{ik} approach the special relativistic values sufficiently fast as $r=((x^1)^2+(x^2)^2+(x^3)^2)^{\frac{1}{2}}\to\infty$. Then the integrals

$$J_i=\int\mathfrak{S}_i^4 d^3x$$

(extended over a three-dimensional section $x^4=\text{const}$) exist, and in virtue of the local conservation laws $\partial\mathfrak{S}_i^j/\partial x^j=0$ they are independent of the "time" x^4.

[13] Explicit expressions for $\mathfrak{U}_i^j$ in reference 2 (bibliography), p. 622, Eq. (185).

It follows from (C) that then

$$J_i = \lim_{a\to\infty}\left(\kappa^{-1}\int_{S_a} \mathfrak{A}_i^{\beta 4} n_\beta d\sigma\right) \quad (\beta=1, 2, 3).$$

(S_a is the "sphere" $r=a$, and n_β are the components of its outward normal.) These last integrals are, in many cases, easy to evaluate, and they show explicitly that total energy and momentum are uniquely determined by the values of the g_{ik} for large r, irrespective of the specific nature of the system considered.[14] Likewise, if we apply a coordinate transformation which does not change the asymptotically Minkowskian character of the metric and is thus asymptotically (for large r) a Lorentz transformation, the J_i themselves undergo this same Lorentz transformation.[15]

2(d) Problem of Motion

Of the later developments in the theory of general relativity proper (i.e., excluding for the present the attempts at constructing a unified field theory) I shall mention only the analysis of the problem of motion, which is of considerable interest.

This problem concerns two related, but separate, questions. The first is the motion of a test particle (a particle of vanishing mass) in a given gravitational field. The second is the n-body problem where the task is to find both the gravitational field produced by n particles and the motion of these particles. I shall briefly review some of the more recent work in this direction.

Motion of a Test Particle

We have seen that the principle of the geodesic line is one of the assumptions of the theory, adopted from special relativity. Rather early the question was raised whether it could be derived from the field equations. Infeld and Schild[16] treat this problem as follows: Given a gravitational field g_{ik} free from singularities. The world line L_0 of a test particle is regarded as a limiting case (for $m\to 0$) of the world line $L_{(m)}$ of a particle of mass m, which is treated as a singularity of the field. Thus, no energy-momentum tensor (T_{ij}) need be introduced, and outside $L_{(m)}$ the free field equations $G_{ij}=0$ (or $R_{ij}=0$) are satisfied. Both the field and the world line $L_{(m)}$ depend on m. As $m\to 0$, it is assumed that the field converges to the given field g_{ik}, and that the world line approaches a limiting position L_0, the test particle's world line. Infeld and Schild show that L_0 is necessarily a *geodesic line* of the metric g_{ik}.

n-Body Problem

This problem, which is considerably more comprehensive and also more difficult, has been investigated by two groups: By Einstein and his co-workers[17] (in particular Infeld), and by Fock[18] and some of his students. Both groups use approximation methods adapted to the assumption that, in a suitable frame of reference, the velocities are small compared to the velocity of light and that the gravitational field is relatively weak (so that the Minkowskian metric may serve as a starting point for the approximation procedure). In addition, both assume an asymptotically Minkowskian metric, hence treat the n bodies as an isolated system.

Einstein treats the n bodies as point particles, i.e., as singularities of the gravitational field and has, therefore, merely to satisfy the free field equations ($R_{ik}=0$) outside the singularities, so that he can avoid the introduction of a specific energy-momentum tensor. In addition, it is explicitly assumed that each singularity has the character of a pole—not that of a dipole or higher pole—and that no spontaneous radiation is emitted. The masses of the particles appear as integration constants, and it follows from the analysis that they are time independent. The first approximation yields the Newtonian potential and the Newtonian law of motion. The second approximation yields additional terms, from which, for example, the advance of the perihelion of a planetary orbit may be derived.

Fock works from the start with extended bodies described by a suitable energy-momentum tensor T_{ij} and deals thus with a combined gravitational and hydrodynamical problem. If the linear dimensions of the n bodies are small compared to their mutual distances, he obtains the same results as Einstein. He has, however, gone further by allowing not only translatory but also rotatory motion, and has also investigated the (gravitational) radiation problem in detail.

2(e) Mathematical Investigations of the Free Field Equations

Since the gravitational field equations are obtained by generalizing Poisson's equation

$$\nabla^2\phi = 4\pi\gamma\rho$$

the energy-momentum tensor replacing the material

---

[14] For a static field whose metric form coincides with the Schwarzschild line element for large r one finds $J_4=mc^2$, $J_i=0 (i\neq 4)$. Here m is the mass constant that appears in the Schwarzschild line element.

[15] The connection between the covariance of the field equations and the identities and conservation laws was recognized very early. The method described here was developed and studied by Einstein, Hilbert, Klein, Lorentz, and others. [For full references see Pauli, art. 23.) But it was not until 1939 that an explicit form for (C) was first derived (by Freud, reference 18 (bibliography).] The particularly simple expressions used here are due to Landau and Lifschitz [reference 5 (bibliography), Eq. (11–78)].—For recent investigations on this subject see references 35a, 19, 20, 33 (bibliography).

[16] See reference 21 (bibliography).

[17] See reference 22 (bibliography).

[18] The whole Chapter VI of Fock's book (reference 6, bibliography) is devoted to the problem of motion.

density ρ, one may expect similar results for them, in particular, that for vanishing T_{ij} no stationary gravitational field exists. I refer, of course, to a field free of singularities. This has been proved under very general assumptions. The strongest results are due to Lichnérowicz.[19] They relate to a field which, in a suitable frame of reference, is independent of the time like variable x^4. The assumptions on the three-dimensional space S of the coordinates x^1, x^2, x^3 are of two kinds: Either S is *closed*, in which case no further conditions are needed; or S is *infinite* (topologically Euclidean) and, in addition, *asymptotically Minkowskian.* In both cases the Riemann curvature tensor vanishes, and the space-time continuum is (at least locally) Minkowskian, although for a closed S it is, of course, topologically different from the Minkowskian world of special relativity.

Very little is known on the nonstationary case. (The classical analog would be the wave equation—instead of Laplace's equation—which does admit solutions with the proper asymptotic behavior.) This question certainly deserves further study. It would be highly interesting to know what kind of solutions, if any, are possible.

In this context I should like to mention Wheeler's work,[20] which deals no longer with the pure gravitational field, but with the Einstein-Maxwell theory, the combined gravitational and electromagnetic equations for a field free of charges and currents. Wheeler constructs solutions, his "geons" (short for gravitational-electromagnetic entities), which have the properties of "bodies": The electromagnetic field remains confined to a finite region under the gravitational attraction produced by its own energy density. The mathematical analysis is partly based on approximations, but it has been made highly probable that rigorous solutions of this kind exist.

Due to the "weakness" of gravitational interaction the ratio of the linear dimension R and the mass M of a geon has the uncomfortably small order of magnitude

$$\gamma/c^2 = 0.74\times 10^{-28}\ \text{cm/g}$$

so that either the extension or the density of a geon is tremendous.

2(f) Verification of General Relativity by Astronomical Observations[21]

There are the three classical astronomical effects. (1) The observations on the advance of the Mercury perihelion are now in very good agreement with the theory. (2) As to the deflection of light in the gravitational field of the sun, it is generally admitted that an effect of the predicted order of magnitude exists, but there is no unanimity among astronomers as to whether the theory is quantitatively confirmed.[22] (3) Finally, there is the gravitational shift of spectral lines. The theory predicts

$$(\nu_0 - \nu_e)/\nu_e = (\phi_e - \phi_0)/c^2.$$

Here, ν_e and ν_0 are emitted and observed frequencies, respectively. The emitting light source and the observer are assumed at rest in a static gravitational field. ϕ_e and ϕ_0 are the corresponding values of the Newtonian potential, which provides an excellent approximation in all cases of interest. For a star, $\phi_e = -\gamma M/R$ (M=mass, R=radius of the star), while the terrestrial ϕ_0 is negligible. Thus we have the red shift

$$(\nu_0 - \nu_e)/\nu_e = -\gamma M/Rc^2$$

which for the sun has the order of magnitude 2×10^{-6}. For rays coming from the limb of the solar disk, the observed red shift has the theoretical value, but it is smaller for all other rays from the sun. Although an explanation for this behavior may be given, this is still an inconclusive result. The theory seems best confirmed for some stars of very high density (e.g., the companion of Sirius), for which the red shift should indeed be considerably bigger than it is for the sun. Unfortunately, in these cases M and R are not very accurately known.

It may be that in the not too distant future terrestrial experiments will, at long last, become possible. For an observer on the earth, and the emitting light source at rest at height h, we have a "violet" shift

$$(\nu_0 - \nu_e)/\nu_e = gh/c^2$$

(g=gravitational acceleration on the earth's surface). If h=1 km, this shift is of the order 10^{-13}, still outside the range of the most accurate detecting devices. Here the artificial satellites[23] may provide the solution in magnifying the effect by a factor 10^3. (The Doppler effect due to the satellite's motion changes the shift into a red shift, as long as h is smaller than half the earth's radius, but does not appreciably alter the order of magnitude of the effect.)

Unfortunately, I do not have the time to talk here about the field of cosmology, the field on which general relativity has had by far the most fruitful and stimulating effect and which is now advancing so rapidly, due to the recent extraordinary achievements of astronomy. Let me stress just one fact. For a long time, as you know, the observed value of the Hubble constant was very disquieting because the age of the universe appeared smaller than the age of the earth's crust as determined from radioactive measurements. During the last few years, however—mainly through Baade's work[24]—the distance scale of distant nebulae has more

[19] See reference 3, Chap. VIII (bibliography).
[20] See reference 23 (bibliography).
[21] For this whole section compare Ginsburg's excellent survey, reference 24 (bibliography).

[22] See R. Trumpler's report to the Berne conference, reference 11 (bibliography). Trumpler considers the observational data in good agreement with the theory.
[23] See references 24 and 25 (bibliography).
[24] See W. Baade's report to the Berne conference, reference 11 (bibliography).

than doubled. According to recent determinations of the Hubble constant, $1/H$ is of the order of 5×10^9 years. If this is confirmed, then the main difficulty will have disappeared.

Outside of cosmology, the impact of general relativity on the rest of physics has not been nearly so great as that of special relativity. One of the reasons is undoubtedly the smallness of the observable effects predicted by the theory and, generally speaking, the weakness of gravitational interactions, particularly in comparison with other interactions between elementary particles. The smallness of direct gravitational effects, however, does not prove that the principles of general relativity are unimportant even for microscopic phenomena. If all frames of reference are equivalent with respect to the laws of macrophysics, it is hard to believe that this equivalence has no counterpart in microphysics, even though the mathematical form in which the equivalence is expressed may greatly differ from the customary one.

To quote Klein[25]: "In atomic physics there has been a tendency to neglect the importance of general relativistic invariance on account of the small influence of gravitational forces on ordinary phenomena. Quite apart from the fact that gravitation will probably be important at very small distances[26] such a point of view would seem no more justified than would be the neglect of atomic spin in atomic structure because in most problems the direct action of spin magnetic moments is very small."

3. UNIFIED FIELD THEORIES

I shall now take up the attempts to go beyond the theory of general relativity in its original form. There are two main trends: on one hand, the search for a unified field theory, on the other, the attempts to quantize the gravitational equations.

Throughout the years many unified theories have been suggested and investigated. It is not my intention to consider them in any detail; I shall rather try to characterize some of the main ideas and procedures which have been developed.

General relativity, as we have seen, leads in an essentially unambiguous way only to a theory of the gravitational field. Accordingly, in the field equations,

$$G_{ij}=-\kappa T_{ij},$$

only the left-hand side is determined. General relativity can accommodate virtually any additional field and its energy-momentum tensor, but it provides no method for choosing among the various possibilities.

When Weyl, in 1918, developed the first unified field theory[27] the unification of various fields of physics had progressed very far. The study of atomic phenomena, in particular, had revealed the predominance of electromagnetic interactions, and it seemed probable that in the last analysis all physical forces could be reduced to electromagnetic and gravitational ones. It was therefore very tempting to complete the synthesis by combining gravitation and electromagnetism in a more comprehensive entity. The unification was to achieve two things: (1) to deduce, at least in principle, all physical interactions from one law, (2) to modify the field equations in such a way that they would admit solutions corresponding to stable charged particles.

In the course of time, specifically after the advent of quantum mechanics, the program which I have sketched has undergone many changes and has been interpreted in many different ways. There was the possibility of interpreting the "particles" either as part of the field or as genuine point singularities; there was the possibility of interpreting a unified field theory either as a c-number theory later to be quantized or as a field theory which implicitly contained the quantum laws already. Also the view that the gravitational and electromagnetic fields were the only fundamental ones was no longer generally accepted.

So far, none of these theories has actually attained the objectives of the original program. To a large extent the work has been concerned with the search for a mathematical structure—more comprehensive than the four dimensional Riemannian space—on which to base a unified or generalized field theory, even a theory that would effect only a partial unification.

While in general relativity the equivalence principle and the principle of general covariance lead directly to the metric structure of space-time and also to the explicit form of the field equations, no such general principles are available to guide the search for a more comprehensive structure. As a result, the search has proceeded in quite different directions.

I shall discuss here only two types of theories, both of which have been actively investigated in recent years.

3(a) Five-Dimensional Theory of Kaluza-Klein[28]

This theory starts from the following observation, which is far from trivial. The fourteen field variables—namely, the ten g_{ik} and the four components ϕ_i of the electromagnetic potential—and the fourteen field equations that describe the gravitational and electromagnetic fields in a four-dimensional space-time continuum S_4 may be interpreted in terms of a suitable *five-dimensional* Riemannian space S_5.

S_5 is characterized by the following property: In an appropriate coordinate system all components $g_{\alpha\beta}$ of

[25] Reference 32, p. 17 (bibliography).

[26] Landau points out that for this reason quantum electrodynamics may not be considered a closed system, because for very high cut-off momenta (i.e., for very small distances) gravitational effects can no longer be neglected. See reference 26, p. 60 (bibliography).

[27] See reference 7 (bibliography). This theory was later abandoned by Weyl and will not be considered here.

[28] The original papers are references 27 and 28 (bibliography).

its metric tensor[29] are *independent* of the fifth coordinate x^5.

Through every point P of S_5 passes a curve $C: x^i = \text{const}$ $(i=1,\cdots,4)$, and any vector a^α at P may be decomposed into two vectors $a_\perp$ and $a_{\parallel}$ which are, respectively, parallel and perpendicular to C. For their lengths $l_{\parallel}$ and $l_\perp$ one obtains

$$l_{\parallel}^2 = (\bar{g}_{\alpha 5} a^\alpha)^2 / \bar{g}_{55};$$
$$l_\perp^2 = \bar{h}_{ij} a^i a^j; \quad \bar{h}_{ij} = \bar{g}_{ij} - \bar{g}_{i5}\bar{g}_{j5}/\bar{g}_{55}.$$

The correspondence of the field variables in S_4 with the metric in S_5 is then defined by

$$\begin{aligned} g_{ij} &= \bar{h}_{ij} \quad (i,j=1,\cdots,4), \\ 2^{\frac{1}{2}}\phi_i &= \bar{g}_{i5}/\bar{g}_{55} \quad (i=1,\cdots,4). \end{aligned} \tag{3.1}$$

This identification[30] is justified by the following facts:

Transformation Properties

There are two types of coordinate transformations in S_5 which leave the curves C unchanged:

$$x^{i'} = F^i(x^1,\cdots,x^4); \quad x^{5'} = x^5; \tag{3.2a}$$

$$x^{i'} = x^i; \quad x^{5'} = x^5 + \zeta(x^1,\cdots,x^4). \tag{3.2b}$$

Under (3.2a), $\bar{h}_{ij}$, ϕ_i, $\bar{g}_{55}$ transform respectively like a tensor, a vector, and a scalar in four-dimensional space. Under (3.2b), $\bar{h}_{ij}$ and $\bar{g}_{55}$ are unchanged, but ϕ_i is subjected to a *gauge transformation*

$$\phi_i' = \phi_i - 2^{-\frac{1}{2}}\partial\zeta/\partial x^i.$$

Field Equations

In the original form of the theory Kaluza added the condition

$$\bar{g}_{55} = -1,$$

thus reducing the number of independent $\bar{g}_{\alpha\beta}$ to fourteen, the number of field variables. Under this assumption one finds for $|\bar{g}|^{\frac{1}{2}}\bar{R}$ ($\bar{R}$ the curvature scalar of S_5, and $\bar{g}$ the determinant of the metric tensor) the expression

$$|\bar{g}|^{\frac{1}{2}}\bar{R} = |g|^{\frac{1}{2}}(R - \tfrac{1}{2}f_{ik}f^{ik}) \tag{3.3}$$

($f_{ik} = \partial\phi_i/\partial x^k - \partial\phi_k/\partial x^i$), i.e., precisely the Lagrangian $\mathfrak{R} - \mathfrak{L}_s$ of the Einstein-Maxwell theory [see Eq. (2.11a)] in units where $\kappa = 1$. The field equations themselves may be expressed by the contracted curvature tensor of S_5. The gravitational equations read

$$\bar{G}^{ij} \equiv G^{ij} + T^{ij} = 0$$

(where T^{ij} is the Maxwellian energy-momentum tensor), and Maxwell's equations are

$$\bar{G}_5{}^i \equiv 2^{-\frac{1}{2}} f^{ij}{}_{;j} = 0.$$

The world line of a charged test particle in four-dimensional space-time corresponds to a geodesic in S_5. Such a geodesic intersects each curve C it meets at a fixed angle which is determined by the specific charge e/m of the test particle.

So far the "five-dimensional" theory merely provides a reinterpretation of the Einstein-Maxwell theory in quasi geometrical terms without changing its content in any way.[31] Nevertheless, the relations discovered by Kaluza and Klein have repeatedly attracted the attention of mathematicians and physicists, and they have suggested genuine generalizations.

The theory of Jordan-Theory.[32] If Kaluza's assumption $\bar{g}_{55} = -1$ is dropped a fifteenth variable, say, $\chi = -\bar{g}_{55}$, which is a scalar in four-dimensional space-time, is introduced into the theory. What is its significance? The expression for the curvature scalar $\bar{R}$ changes from (3.3) to

$$\bar{R} = R - \tfrac{1}{2}\chi f_{ik}f^{ik} + g^{ik}(\chi_i/\chi)_{;k}. \tag{3.4}$$

($\chi_i = \partial\chi/\partial x^i$), which suggests the interpretation of χ as a *variable* gravitational constant. On this basis Jordan has developed an extensive theory. It is based on a variational principle whose Lagrangian is slightly more general than $|\bar{g}|^{\frac{1}{2}}\bar{R}$ and contains two adjustable constant parameters. In addition to the fourteen Einstein-Maxwell equations—which are, of course, modified—there appears a fifteenth field equation for χ.

Dirac[33] first advanced the idea of a variable gravitational constant. He argued, roughly, as follows. For two electrons the ratio of the Coulomb force and the Newtonian gravitational force is the dimensionless number $e^2/\gamma m^2$ of the order 10^{42}. It is highly unlikely that a constant of this size may be explained by any theory.

On the other hand, numbers of this order are obtained if astronomical magnitudes are expressed in atomic units (for example, m_0 = proton mass, l_0 = electron radius, $t_0 = l_0/c$). Thus, we find $\sim 10^{40}$ for the present age of the universe (of some billion years), $\sim 10^{-40}$ for the average material density in the universe, etc. The difficulty of explaining the value of $e^2/\gamma m^2$ disappears, therefore, if γ is not a constant but inversely proportional to the age of the universe, while the atomic units remain unchanged. This argues for a time dependence of the gravitational constant, and in a relativistic theory γ must then be treated as a function of space-time.

In a very interesting critical analysis Fierz[34] has pointed out that Jordan's theory in its present form (which includes only the pure electromagnetic field) does not yet lead to an unambiguous interpretation of

---

[29] In Sec. 3(a), barred quantities refer to the metric in S_5. Greek indices run from 1 to 5, Latin indices from 1 to 4.

[30] Kaluza identified the metric of S_4 with $\bar{g}_{ij}$ $(i,j=1,\ldots,4)$ and therefore obtained the Einstein-Maxwell equations only as a first approximation for weak fields. The equations (3.1) are due to Klein.—The sign of $\bar{g}_{55}$ must be so chosen that the curves C are space-like.

[31] This holds also for the mathematically more elegant versions of the theory. The "projective" theory is described in reference 8 (bibliography), Chap. III.

[32] See Jordan's book [reference 8 (bibliography)] and reference 29 (bibliography).

[33] See reference 30 (bibliography).

[34] See reference 31 (bibliography).

either χ as gravitational constant or g_{ij} as the true metric tensor. For an unambiguous interpretation it would be necessary to make specific assumptions about the action on material particles. Depending on these assumptions also the ratios of different atomic units of length may or may not be constant.

Apart from any particular theoretical views it would be very worthwhile to attack the question of the time dependence of γ by experiment comparing "astronomical" and "atomic" times.[35]

Klein's theory.[36]—O. Klein has generalized the original Kaluza-Klein theory in an entirely different way. He drops the condition that the field quantities be independent of x^5, and develops a truly five-dimensional theory. The fifth coordinate has a quantum theoretical significance.

As I have mentioned, the world lines of charged particles correspond to geodesics in Kaluza's S_5. These may be obtained from the Lagrangian

$$L=\bar{g}_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta \quad (\dot{x}^\alpha=dx^\alpha/d\tau)$$

(τ is an auxiliary parameter), and in suitable units the specific charge e/m of the particle equals the momentum $p_5=\partial L/\partial\dot{x}^5$. The fifth coordinate is thus conjugate to the charge, and in view of the quantization of charge Klein assumes x^5 to be an angular variable, or, in other words, he assumes the field variables to be periodic in x^5. Expanding each field variable in a Fourier series in x^5, one obtains a sequence of Fourier coefficients which are functions of the space-time coordinate x^i. They presumably refer to different particles and are to be quantized. Particles with half-integral spin may also be incorporated in the theory.

3(b) Nonsymmetric Theories of Einstein and Schrödinger[37]

In contrast to the theory of Kaluza and Klein, these theories operate in the ordinary four-dimensional space-time continuum.

While the basic element of the mathematical structure of general relativity is the Riemannian metric ds^2 of the space-time continuum, the mathematical reinvestigation of infinitesimal geometry—which was greatly stimulated by Einstein's theory—has revealed the more fundamental structural element of the affine connection (or "affinity") or the parallel displacement of vectors defined by the equations

$$\delta a^i=-\Gamma_{jk}{}^i a^j dx^k.$$

In the hands of Levi-Cività and of his successors this new concept has, indeed, led to a very remarkable mathematical progress. It is more fundamental because it remains applicable if no metric is defined. The curvature tensor, in particular, is directly defined in terms of the Γ's.

Quite generally one finds that the symmetric part

$$^s\Gamma_{jk}{}^i=\tfrac{1}{2}(\Gamma_{jk}{}^i+\Gamma_{kj}{}^i) \qquad (3.5a)$$

of an affinity is itself an affinity, while the antisymmetric part

$$^a\Gamma_{jk}{}^i=\tfrac{1}{2}(\Gamma_{jk}{}^i-\Gamma_{kj}{}^i) \qquad (3.5b)$$

is a tensor. Accordingly, the symmetric and antisymmetric parts transform separately. In a Riemannian space an affinity is uniquely defined by the requirement that it be symmetric and that the length of a vector remain unchanged by parallel displacement [see Eq. (2.4)].

The first who based a unified field theory on the affinity—instead of the metric—was Eddington. (The affinity was assumed *symmetric.*) Both the metric and the electromagnetic field appeared as derived quantities. Eddington did not postulate specific field equations. This was done, in 1923, by Einstein, who derived the field equations from a Lagrangian depending only on the components R_{ik} of the contracted curvature tensor. Soon after, however, Einstein abandoned the theory because the electromagnetic field equations appeared to lead to inadmissible consequences (for example, the vector potential was proportional to the current).

For many years this theory lay dormant until it was revived, about ten years ago, by Einstein and his co-workers and by Schrödinger. The most significant change was the introduction of a nonsymmetric affinity and of a nonsymmetric metric tensor. Schrödinger introduced only the Γ's as primary field variables (obtaining the g_{ij} as derived quantities); Einstein, on the other hand, used both the $\Gamma_{jk}{}^i$ and the nonsymmetric $\mathfrak{g}^{ij}$ as building stones for the theory.

For a nonsymmetric g_{ij} the equations

$$g_{ij}g^{kj}=\delta_i{}^k; \quad \mathfrak{g}^{ij}=|g|^{\frac{1}{2}}g^{ij}$$

define g^{ij} and $\mathfrak{g}^{ij}$. The symmetric and antisymmetric parts of $\mathfrak{g}^{ij}$ are given by

$$^s\mathfrak{g}^{ij}=\tfrac{1}{2}(\mathfrak{g}^{ij}+\mathfrak{g}^{ji}); \quad ^a\mathfrak{g}^{ij}=\tfrac{1}{2}(\mathfrak{g}^{ij}-\mathfrak{g}^{ji}). \qquad (3.6)$$

In the following, I shall mainly report on Einstein's form of the theory. Since the symmetric and antisymmetric parts of the Γ and of the $\mathfrak{g}^{ij}$ transform separately under coordinate transformations the objection might be raised that no genuine unification is achieved, and that the field equations will not be unambiguously defined by the covariance group of the theory. To the extent that it applies to the affinity Einstein has met this objection in the following way.

Invariance under λ transformations.—It is postulated

[35] The feasibility of such an experiment was discussed by G. M. Clemence and R. H. Dicke at an astrophysical seminar at Princeton (April, 1955).

[36] See reference 32 (bibliography) and Klein's report to the Berne conference.

[37] For the final version of Einstein's theory see reference 33 (bibliography), and "Meaning of Relativity," Appendix II. For Schrödinger's theory see reference 34 (bibliography) and his book, reference 9 (bibliography). An excellent account of the nonsymmetric theories is given in M. A. Tonnelat's book [reference 10 (bibliography)], which contains also an extensive bibliography.

that two affinities Γ and Γ' describe the same field if

$$\Gamma_{jk}{}^{i\prime} = \Gamma_{jk}{}^{i} + \delta_j{}^i \frac{\partial \lambda}{\partial x^k}, \tag{3.7}$$

where λ is an arbitrary coordinate function. This λ transformation, which resembles the gauge transformation of electromagnetism, mixes the symmetric and antisymmetric parts of $\Gamma_{jk}{}^i$; these lose therefore their independent character if the λ transformations are combined with ordinary coordinate transformations to form the full covariance group of the theory. The symmetric and antisymmetric parts of $\mathfrak{g}^{ij}$, however, remain independent.

The λ transformation has a simple geometrical significance. Under two affinities, Γ and Γ', the ratios $a^i : a^j$ of the components of a vector change in exactly the same way if and only if

$$\Gamma_{jk}{}^{i\prime} = \Gamma_{jk}{}^{i} + \delta_j{}^i \zeta_k,$$

where ζ_k is an arbitrary covariant vector. If we require, in addition, that under Γ and Γ' parallel displacement along *closed paths* is the same for the vector components themselves (not only for their ratios) we find as a necessary and sufficient condition that ζ_k is a gradient, i.e., $\zeta_k = \partial\lambda/\partial x^k$, as in (3.7).

Since the curvature tensor measures the change of a vector by parallel displacement along infinitesimal closed paths it follows immediately that the curvature tensor as well as its contractions are unaffected by λ transformations.

In order to facilitate the formulation of the theory Einstein and Kaufman replaced the $\Gamma_{jk}{}^i$ by the combinations

$$U_{jk}{}^i = \Gamma_{jk}{}^i - \delta_k{}^i \Gamma_{jt}{}^t$$

which, by a λ transformation, change into

$$U_{jk}{}^{i\prime} = U_{jk}{}^i + \delta_j{}^i \frac{\partial \lambda}{\partial x_k} - \delta_k{}^i \frac{\partial \lambda}{\partial x^j}.$$

No direct geometrical interpretation has been given for them.

Transposition invariance.—The field equations are to be derived from a variational principle

$$\delta \int \mathfrak{L} d^4x = 0.$$

The Lagrangian is a scalar density *invariant under λ transformations*, which is a function of the $\mathfrak{g}^{ij}$ and the $U_{jk}{}^i$. The $\mathfrak{g}^{ij}$ and $U_{jk}{}^i$ are to be varied independently.

In order to narrow down the choice of the Lagrangian, Einstein introduced the postulate of transposition invariance: $\mathfrak{L}$ remains unchanged if $\mathfrak{g}^{ij}$ and $U_{jk}{}^i$ are replaced, respectively, by their transposed

$$\tilde{\mathfrak{g}}^{ij} = \mathfrak{g}^{ji}; \quad \tilde{U}_{jk}{}^i = U_{kj}{}^i.$$

Presumably, this expresses the symmetry of the theory in positive and negative electric charges. The transposition invariance of $\mathfrak{L}$ implies that the field equations admit, together with a field $\mathfrak{g}, U$ also the transposed field $\tilde{\mathfrak{g}}$, $\tilde{U}$ as a solution.

Einstein's choice for the Lagrangian is

$$\mathfrak{L} = \mathfrak{g}^{ij} R_{ij}.$$

It satisfies the conditions of transposition and of λ invariance, and it is a natural generalization of the Lagrangian of general relativity to the nonsymmetric case.

The field equations.—Variation of the $\mathfrak{g}^{ij}$ and of the $U_{jk}{}^i$ yields the two sets of field equations

$$\text{(A)} \quad R_{ij} = 0$$

$$\text{(B)} \quad \frac{\partial \mathfrak{g}^{ij}}{\partial x^k} + \Gamma_{sk}{}^i \mathfrak{g}^{sj} + \Gamma_{ks}{}^j \mathfrak{g}^{is} - \Gamma_{ks}{}^s \mathfrak{g}^{ij} + \tfrac{2}{3}\delta_k{}^j \Gamma_{st}{}^t \mathfrak{g}^{is} = 0.$$

(A) consists of 16 equations. (B) consists of 64 equations, which immediately imply

$$\partial\, {}^a\mathfrak{g}^{ik}/\partial x^k = 0. \tag{3.8}$$

Einstein's original equations for a pure gravitational field ($T_{ij} = 0$) follow from (A) and (B) for symmetric $\mathfrak{g}$ and Γ. The set (B) is then equivalent to Eq. (2.4), and the set (A) reduces to ten equations, because R_{ij} is then symmetric.

Comparison with Schrödinger's theory.—In Schrödinger's theory the primary field variables are the $\Gamma_{jk}{}^i$, the Lagrangian $\mathfrak{L}$ is a function of the R_{ij}, and the $\mathfrak{g}^{ij}$ are defined as

$$\mathfrak{g}^{ij} = \partial \mathfrak{L}/\partial R_{ij}.$$

Schrödinger chooses $\mathfrak{L} = (2/\Lambda)(-\Delta)^{\frac{1}{2}}$. Λ is a constant, and Δ is the determinant of the R_{ij}. The resulting field equations are similar to Einstein's. The set (B) is unchanged, and (A) is replaced by

$$\text{(A}'\text{)} \quad R_{ij} = \Lambda g_{ij}$$

so that Λ plays the role of a cosmological constant.

Content of the theory.—While the mathematical basis of the theory has been perfected, and may now be presented in a simple and transparent form, very little can as yet be said about the content of the theory. Even the interpretation of the field variables is difficult; it cannot be totally separated from the analysis of the field equations and of their solutions.[38]

A few clues, however, are available regarding the symmetric and antisymmetric parts of the tensor $\mathfrak{g}^{ij}$. The first is furnished by the equations (3.8). These coincide with the second set of Maxwell's equations if we put

$$\mathbf{E} = \alpha({}^a\mathfrak{g}^{23}, {}^a\mathfrak{g}^{31}, {}^a\mathfrak{g}^{12}); \quad \mathbf{B} = \alpha({}^a\mathfrak{g}^{14}, {}^a\mathfrak{g}^{24}, {}^a\mathfrak{g}^{34})$$

[38] The assumption II of Sec. 2(a) is no longer applicable, and presumably the gravitational and electromagnetic fields, for example, can only be clearly separated in the limit of very weak fields.

with some constant α. Hence, Einstein suggests identifying ${}^a\mathfrak{g}^{ij}$ with the electromagnetic field.

Secondly, Lichnérowicz has shown what corresponds to the local light cone of relativity, as far as the propagation of waves (or disturbances) is concerned.[39] By an analysis of the continuation properties of the field equations he finds

$$l_{ij}dx^i dx^j = 0,$$

where ${}^s\mathfrak{g}^{kj}l_{ij} = \delta_i{}^k$. Note that, in general, l_{ij} is not proportional to the symmetric part of the covariant tensor g_{ij}. Note also that the light cone defines only the ratios of the l_{ij} and hence only the ratios of the ${}^s\mathfrak{g}^{ij}$.

Being derived from an invariant variational principle the field equations satisfy the necessary identities and lead to conservation laws [see Sec. 2(c)]. Because of prohibitive mathematical difficulties, it has not been possible to derive much more than that from the field equations.

A number of explicit solutions (with special symmetries) have been found, but all of them are singular. According to Einstein's program, however, the field equations (A) and (B) must be considered complete, and their solutions must be *free of singularities*. So far no such solution (for reasonable boundary conditions) has been found, nor has it been shown that no such solution exists.

Einstein has made it quite clear that he regarded the variables of this theory as classical field variables which were not to be quantized.

4. REMARKS ON THE QUANTIZATION OF COVARIANT THEORIES

There have been, of late, a number of investigations on the quantization of covariant field theories. The most extensive work has probably been done by P. G. Bergmann and his students.[40] For details I refer to his report to the Berne conference.

Since the mathematical analysis is rather complex, I shall be quite brief, and shall confine myself to pointing out the specific difficulties which have their root in the covariance of the theory to be quantized.

In applying the Heisenberg-Pauli method of field quantization the first step—still on the c-number level—consists in transforming a theory from its Lagrangian to a Hamiltonian form. This leads to difficulties whenever the momenta are not independent. The best known example is Maxwell's theory, where π^4, the momentum conjugate to the scalar potential, vanishes identically.

Any covariant theory leads to four independent relations between the momenta, the "primary" constraints in Bergmann's terminology. These are obtained by the same method which furnishes the identities and conservation laws for a covariant field theory and which has been sketched in 2(c).[41] The condition that the primary constraints, if satisfied at some initial time, remain valid in virtue of the Hamiltonian equations of motion yields four additional, the "secondary" constraints.

The situation has been sufficiently well analyzed, so that it is known, at least in principle, how to proceed. The difficulty remains, however, to express the relations which have been derived—notably for the gravitational field—in a tractable form.

There still remains the possibility of by-passing the Hamiltonian formalism and of working, instead, directly with the Lagrangian, in analogy to Feynman's method of quantization. Attempts in this direction are being made.

The experience with quantum electrodynamics would seem to indicate that, in any event, a thorough knowledge of the solutions of the classical field equations is indispensable, which makes the investigation of non-stationary solutions (mentioned in 2e) all the more important.

Since the work, so far, has been preparatory to an actual quantization it is not known to what extent specifically quantum theoretical difficulties may arise.

REMARK BY H. P. ROBERTSON IN THE DISCUSSION

I would like to know if any of you or your experimental colleagues know of any attempts at the repetition of the Kennedy-Thorndike experiment. It seems to me that the time may be ripe for the retest of this experiment which is radically different in principle from the Michelson-Morley. I think the other two experiments are adequate.

BIBLIOGRAPHY

I. Books

1. Albert Einstein, *Meaning of Relativity* (Princeton University Press, Princeton, 1955), fifth edition.
2. W. Pauli, *Encyklopädie der Mathematischen Wissenschaften* (B. G. Teubner, Leipzig, 1921), Vol. V2, pp. 539–775.
3. A. Lichnérowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Masson et Cie, Paris, 1955).
4. P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Inc., New York, 1942).
5. L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Press, Cambridge, 1951), English translation.
6. V. A. Fock, *Theory of Space, Time and Gravitation* (Gosisdat, Moscow, 1955), in Russian.

[39] See reference 3, p. 288.

[40] See reference 35 (bibliography).

[41] Rosenfeld [reference 36 (bibliography)] has already shown that relations of this kind exist if the Lagrangian of the theory admits a group of transformations which depend on arbitrary space-time functions [such as the four functions ξ^i in 2(c) which determine the infinitesimal transformations of the group]. In Maxwell's theory the corresponding group is the gauge group. It involves only one arbitrary space-time function and leads to only one relation, *viz.*, $\pi^4 = 0$.

7. H. Weyl, *Raum. Zeit. Materie* (Verlag Julius Springer, Berlin, 1923), fifth edition.
8. P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1955), second edition.
9. E. Schrödinger, *Space-Time Structure* (Cambridge University Press, Cambridge, 1950).
10. M. A. Tonnelat, *La thèorie du champ unifié d'Einstein* (Gauthier-Villars, Paris, 1955).
11. Proceedings of the Conference on Relativity at Berne, 1955. To appear shortly as a supplement to Helv. Phys. Acta.

II. Papers

12. H. P. Robertson, Revs. Modern Phys. **21**, 378 (1949).
13. Durbin, Loar, and Havens, Phys. Rev. **88**, 179 (1952).
14. W. Pauli, Phys. Rev. **58**, 716 (1940). Also Progr. Theoret. Phys. **5**, 527 (1950).
15. E. P. Wigner, Ann. Math. **40**, 149 (1939).
16. (a) V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. **34**, 211 (1948); (b) T. D. Newton and E. P. Wigner, Revs. Modern Phys. **21**, 400 (1949).
17. C. Møller, Dan. Mat. Fys. Medd. **30**, No. 10 (1955).
18. Ph. Freud, Ann. Math. **40**, 417 (1939).
19. A. Einstein, Can. J. Math. **2**, 120 (1950).
20. E. Schrödinger, Proc. Irish Acad. **54A**, 79 (1951).
21. L. Infeld and A. Schild, Revs. Modern Phys. **21**, 408 (1949).
22. (a) Einstein, Infeld, and Hoffmann, Ann. Math. **39**, 66 (1938); (b) A. Einstein and L. Infeld, Ann. Math. **41**, 455 (1940); (c) A. Einstein and L. Infeld, Can. J. Math. **1**, 209 (1949); (d) L. Infeld, Acta Phys. Polonica **XIII**, 187 (1954).
23. J. A. Wheeler, Phys. Rev. **97**, 511 (1955).
24. V. L. Ginsburg, Uspekhi fisz. Nauk, **LIX**, 11–49 (May, 1956).
25. S. F. Singer, Phys. Rev. **104**, 11 (1956).
26. L. D. Landau, "On the quantum theory of fields," in *Niels Bohr and the Development of Physics* (New York and London, 1955).
27. Th. Kaluza, Sitz. Preuss. Akad. Wiss. 966, 1921.
28. O. Klein, Z. Physik **37**, 895 (1926).
29. Y. Thiry, Compt. rend. acad. sci. (Paris) **226**, 216 (1948).
30. P. A. M. Dirac, Proc. Roy. Soc. (London) **A165**, 199 (1938).
31. M. Fierz, Helv. Phys. Acta **XXIX**, 128 (1956).
32. O. Klein, Arkiv Mat. Astron. Fysik B **34A**, No. 1 (1946).
33. A. Einstein and B. Kaufman, Ann. Math. **62**, 128 (1955).
34. E. Schrödinger, Proc. Roy. Irish Acad. **51**, 163 (1947); **51**, 205 (1948); **52**, 1 (1948).
35. (a) P. G. Bergmann, Phys. Rev. **75**, 680 (1949); (b) P. G. Bergmann and J. H. M. Brunings, Revs. Modern Phys. **21**, 480 (1949); (c) J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951); (d) P. G. Bergmann and I. Goldberg, Phys. Rev. **98**, 531 (1955).
36. L. Rosenfeld, Ann. inst. Henri Poincaré **2**, 25 (1932).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 2 APRIL, 1957

Nuclear Physics

VICTOR F. WEISSKOPF

Department of Physics and Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts

I

WITHIN the framework of a short review it is impossible to discuss all outstanding problems in nuclear physics. I therefore restrict myself to the problem of nuclear structure and stability. The main problem can be formulated as follows: Given A nucleons, what can we conclude from our present incomplete knowledge of nucleon interactions, in regard to the structure and properties of the nucleus which they form?

Our knowledge of interactions is, in fact, extremely limited. The properties of the deuteron and the nucleon-nucleon scattering experiments contain much information about the interaction between nuclear pairs. We have deduced phase shifts as functions of energy from many scattering experiments, but so far it was not possible to explain these phase shifts as arising from a well-defined potential. Hence, it is hard to apply this knowledge to the situation within nuclei which is so different from the scattering of free and isolated nucleons. Because of the proximity of many scattering centers within one wavelength inside the nucleus, the conditions of scattering are fundamentally altered; any definite conclusions from the two-body scattering can be drawn only with a knowledge of the scattering potential and not from the mere phase shifts. As an example, let us look at a slow nucleon whose wavelength $\lambda\!\!\!^{-}$ is large compared to the radius r^* of the scatterer. The scattered wave ψ_{sc} will then be an S wave ($\psi_{sc}=[\sin(kr-\delta)]/r$). If it hits a second scatterer at a long distance $d\gg\lambda\!\!\!^{-}$ from the first scatterer, the wave will be essentially constant over the second scattering region, and there will again be mainly S scattering. However if the second scatterer is near the first ($d\ll\lambda\!\!\!^{-}$), the scattered wave ψ_{sc} is far from constant over the scattering region because of the preponderance of the $1/r$ dependence, if $d\ll\lambda\!\!\!^{-}$. In this case, p scattering will be quite important in spite of the low energy. In the usual mathematical terminology, this condition is expressed by saying that the elements of the scattering matrix between states of different energy are also important in our problem. This makes our problem rather difficult since the scattering phase shifts only determine the scattering matrix elements between states of equal energy.

Apart from the incomplete knowledge of the interaction between two free nucleons, our problem is complicated by the fact that we do not know whether this interaction is changed considerably when the two nucleons are surrounded by other nucleons. The

existence of such a many-body effect is neither proved nor disproved. The average distance d of nucleons in nuclear matter, $d \approx 1.8 \times 10^{-13}$ cm, is not very much larger than twice the root mean square r^* of the spacial extension of the proton as measured by the Stanford group, $r^* \sim 0.8 \times 10^{-13}$ cm. Hence, the "meson clouds" of the nucleons almost touch one another in nuclear matter. It is a matter of personal inclination whether the fact $d > 2r^*$ can be used as an argument in favor of a nondense packing in which the forces are expected to be similar to the ones between two isolated particles or whether $d \sim 2r^*$ is an argument in favor of dense packing with appreciable changes in the meson clouds and therefore in the nuclear forces within nuclear matter. At present no direct experimental decision of this alternative has yet been devised.

Since it is not possible to derive the properties of nuclei deductively from the nucleon interactions, we must start at the other end and study the observed properties and try to find some explanations of the salient features of the observed facts. They are:

(A) *The existence of "nuclear matter."* Nuclear shape measurements have shown that all nuclei except the very lightest reach a universal density near the center of about $\rho_0 = 1.7 \times 10^{38}$ particles/cm^3. This seems to be the density at which nucleons reach a certain equilibrium. We are allowed to extrapolate and to assume that, were it not for the Coulomb repulsion, nuclear matter would be thinkable in a stable state of infinite extension and of the density ρ_0. It would consist of an equal mixture of protons and neutrons, having a binding energy of roughly 15 Mev per particle. Here the "saturation" of nuclear forces enters, which prevents nuclear matter both from collapsing to a density less than ρ_0 and from flying apart.

(B) *Independent particle aspect.* The dynamics of nuclear matter exhibits certain unexpectedly simple properties at low excitation energies: They can be fairly well reproduced by assuming that the nucleons within nuclear matter move almost independently of each other in a constant potential. Considering a nucleus as a finite spherical chunk of nuclear matter, we obtain a model of the nucleus consisting of independent nucleons moving within a spherical potential well in the lowest quantum states permitted by the exclusion principle. With the additional assumption of a strong spin-orbit coupling, this model reproduces a surprisingly large amount of experimental facts concerning the properties of nuclei at low excitation. On that basis we can understand the systematics of nuclear binding energies, most of the structure of nuclear spectra, and a good part of the electric and magnetic properties. We learn from this that nuclear matter at low excitation behaves somewhat like a degenerate gas, a fact which poses a grave problem of understanding at the present stage of our knowledge. The source of the spin-orbit coupling is also a problem which has not been satisfactorily explained.

(C) *The third group of important nuclear properties waiting for explanation contains features of less general character.* We quote here the deviation from sphericity found in many nuclei, the problems connected with the interpretation of the ensuing rotational spectra, and the vast amount of material accumulated in the study of nuclear reactions.

As mentioned earlier, there are two approaches to the problem of understanding the principles underlying the nuclear structure. We characterize them by the assumptions $r^* \ll d/2$ and $r^* \gg d/2$, where r^* is the radius of the meson cloud around a nucleon and $\hat{a}$ is the average distance apart of the nucleons. Actually, of course, $r^* \sim d/2$. In the first approach one endeavors to understand the phenomena on the basis of two-body forces between pairs of particles. Examples of this approach are the papers by Brueckner, Eden, Levinson, Mahmoud, Bethe, Goldstone and others.[1] We will call it the Brueckner approach. In the other approach one assumes that the meson clouds of all nucleons merge completely into a one-meson field within the nucleus.

The consequences of the latter approach are difficult to argue in view of our lack of any specific meson theory. It is plausible, however, that such a model would lead to a lack or to a strong reduction of all interactions in nuclear matter at its normal density at which the meson field probably reaches a certain saturation value. This picture has been discussed by Duerr and by Johnson and Teller[2] and contains some interesting speculations as to the nuclear interactions of antinucleons. The "Teller approach" contains the lack of interactions as a basic assumption and explains automatically our group B of facts. There exist certain observations, however, which seem to indicate that the lack of interaction between nucleons in nuclear matter is restricted only to low excitations. If this is correct, this lack is only apparent and is caused by the special properties of nuclear matter in its lowest states, as it would be from the point of view of the Brueckner approach. Some of these observations are: The analysis of nuclear reactions has shown that the mean free path of a nucleon entering nuclear matter is large only for relatively low energy of the particle. It can only be considered as moving independently of the others when the incident energy is less than, say, 10 Mev. At higher energies, but not at too high ones, say from 10 to 50 Mev, the mean free path is quite short compared to the nuclear radius, i.e., it is subject to strong interactions. Other indications of strong forces within the nucleus come from photoproduction of π mesons, from p-d processes, and other pickup reactions. In all these cases, the results indicate that the nuclear wave functions contain momentum components which are much higher than the ones one would get in the case of independent free motion. Hence, strong forces must be present which cause them. How-

[1] For references see H. A. Bethe, Phys. Rev. **103**, 1353 (1956).
[2] H. Duerr, Phys. Rev. **103**, 469 (1956); M. H. Johnson and E. Teller, Phys. Rev. **98**, 783 (1955).

ever the arguments presented here against the Teller approach are not quite convincing. After all, it is only assumed that the interactions disappear in the center of the nucleus where the density reaches the maximum value. The nuclear forces would be functions of the density and assume their known strength at zero density. Hence there would be increasingly stronger interaction towards the surface of the nucleus where the density drops to zero. It is perfectly possible that the effects mentioned before can all be explained by these surface effects. In particular, the π-meson production and the pickup reactions are typical surface phenomena.

The Teller approach will not be further discussed in this review. We would like to face the question whether one can understand the nuclear properties if the forces between the constituents are roughly what we know from the two-nucleon interactions. A large amount of work has been spent recently on this problem by Brueckner and his colleagues, and a formalism has been developed which seems to be adequate to describe the situation. It is a generalization of a formalism originally invented to deal with multiple-scattering phenomena.

There certainly is a good deal of similarity in the problem of a wave scattered by a dense array of centers and the problem of the motion of a nucleon in a nucleus. The path of a nucleon should suffer many scatterings in traversing nuclear matter. The aforementioned group B of facts, however, indicates that there must be a fundamental difference in the two problems which prevents those scatterings from being effective in the nucleus. It was pointed out some time ago[3] that the Pauli exclusion principle plays an important role in this since, from a naive point of view, all end states into which a scattering of two nucleons can lead are already occupied at low excitations. It is not clear, in the first place, why one can use free particle states as actually existing in nuclear matter. However, the present theoretical treatments of the situation seem to bear out this idea and point towards the decisive role of the exclusion principle. Unfortunately, the mathematical difficulties are still very great, and it is not clearly understood what properties of the nuclear force make the independent particle approach a good approximation and lead to the observed density in nuclear matter.

II

Let us now approach the problem in a less deductive way. We learn from experience that the independent particle model is a good description of the facts and we ask what consequences we can draw from this circumstance. It means that the wave function Ψ of the nucleus is approximately a product of one-particle wave functions φ_i, duly antisymmetrized so as to fulfill the Pauli principle:

$$\Psi = A \prod_\alpha \varphi_\alpha(r_i). \tag{1}$$

Here r_i is the coordinate of the ith particle, α is the quantum state in which it is located, and "A" symbolizes the antisymmetrization which makes all α's different for particles of the same kind. The wave functions φ_α are given by a one-particle Hamiltonian

$$H_i\varphi_\alpha = [T_i + U_\alpha(r_i)]\varphi_\alpha = E_\alpha\varphi_\alpha, \tag{2}$$

where the first term is the kinetic energy $T_i = p_i^2/2m$, and the second term is the potential energy, which is assumed to be a potential well of the size of the nucleus. It also contains a spin-orbit term which we will neglect in the following rough estimates. We put approximately

$$\begin{aligned} U_\alpha(r) &= -V_\alpha \quad \text{for} \quad r<R \\ U_\alpha(r) &= 0 \qquad \text{for} \quad r>R, \end{aligned} \tag{3}$$

where R is the nuclear radius and V_α is a positive energy (the well depth). We do not assume that all particles have the same well depth; hence $U_\alpha(r)$ and V_α may depend upon the state α in which the particle is found. The actual exact Hamiltonian can be written in the form

$$H = \sum_{i=1}^{A} T_i + \tfrac{1}{2} \sum_{i,k=1}^{A} V_{ik}(\mathbf{r}_i - \mathbf{r}_k), \tag{4}$$

where V_{ik} is the interaction potential between the pair of particles i and k. We assume that only pair forces exist. If (1) is an approximate solution of (4), then the one-particle potential energy $U_\alpha(r_i)$ is the average effect of all other particles on the ith particle.

The total energy E of the system is given by the expectation value of H in the state Ψ:

$$E = \int \Psi^* H \Psi dr = \sum_\alpha \langle T_\alpha \rangle + \tfrac{1}{2} \sum_\alpha \langle U_\alpha \rangle,$$

where T_α is the expectation value of the kinetic energy in the state α, and $\langle U_\alpha \rangle$ is the average value of U_α in that state α. The sums are extended over all occupied states α. Here the energy appears as a sum of contributions from each occupied state. The factor $\frac{1}{2}$ in the second term is of great importance. It comes from the fact that the potential energy of the particle in the state α is the sum of all pair interactions with the other particles. Hence, if no factor $\frac{1}{2}$ were present, the interaction between a given pair i-k of particles would have been counted twice, once in $\langle U_\alpha \rangle$ and then again in $\langle U_\beta \rangle$, when α is the state of the ith particle and β the state of the kth one. This factor $\frac{1}{2}$ also makes it impossible to define a "model Hamiltonian," i.e., a Hamiltonian for our wave function (1). The Hamiltonian $H = \sum_i H_i$, whose eigenfunctions are the ones defined by (1), has eigenvalues different from (4); namely, the values one would get without the factor $\frac{1}{2}$.

We have assumed that the potentials U_α are square wells with a depth V_α. Hence, neglecting surface effects, each wave function φ_α has a well-defined momentum p_α and a kinetic energy T_α:

$$T_\alpha = p_\alpha^2/2m, \quad \langle U_\alpha \rangle = -V_\alpha.$$

[3] V. Weisskopf, Science 113, 1 (1951).

Also, in the lowest energy state, all levels up to the Fermi kinetic energy T_F are occupied. Then we get

$$\sum_\alpha \langle T_\alpha \rangle = \tfrac{3}{5} A T_F, \tag{5}$$

where the sum is taken over all occupied levels. The packing fraction P is given by

$$P = -\frac{E}{A} = -\tfrac{3}{5}T_F + \tfrac{1}{2}V_{Av}, \quad V_{Av} = \frac{1}{A}\sum_\alpha V_\alpha, \tag{6}$$

where V_{Av} is the average value of the well depth over all occupied states.

We now determine the separation energy S, which is the minimum energy necessary to remove one particle from the nucleus. The easiest one to remove is the one on the top of the Fermi distribution. We get

$$S = -T_F + V_F, \tag{7}$$

where V_F is the well depth for the top particle.[4] Now, S and P are closely related; in fact, in the limit of large nuclei they become equal if surface effects are neglected. The equality of these two magnitudes is an expression of the fact that the total energy is proportional to the number of constituents. By equating P and S we get from (6) and (7)

$$S = -\tfrac{1}{5}T_F + V_{Av} - V_F. \tag{8}$$

This equation demonstrates the necessity of assuming well depths depending on the momentum p_i. If all depths were equal, we would have $V_{Av} = V_F$, and we would get the nonsensical result of a negative separation energy.

Equation (8) does not tell us in what way V_i depends on the state i. Let us make the simplest assumption of a quadratic dependence upon the mementum p_i:

$$V_i = V_0 - \frac{p_i^2}{p_F^2}V_1, \tag{9}$$

where V_0 and V_1 are constants and p_F is the Fermi momentum. Then, of course, $V_{Av} = V_0 - \tfrac{3}{5}V_1$, $V_F = V_0 - V_1$, and we calculate V_0 and V_1 from (7) and (8):

$$V_1 = \tfrac{1}{2}(5S + T_F), \quad V_0 = \tfrac{1}{2}(7S + 3T_F). \tag{10}$$

A quadratic dependence of the well depth can always be expressed in terms of an effective mass m^*. The energy of the ith particle is then written in the form

$$E_i = \frac{p_i^2}{2m} - V_i = \frac{p_i^2}{2m^*} - V_0$$

with

$$\frac{m}{m^*} = 1 + \frac{V_1}{T_F} = \tfrac{3}{2} + \tfrac{5}{2}\frac{S}{T_F}. \tag{11}$$

These relations should only serve as a first orientation and must not be regarded as quantitative. Nevertheless, it is interesting to put in some numbers. The Fermi energy T_F depends only upon the nuclear density and is (with $R = 1.2 \times A^{1/3} \times 10^{-13}$ cm) $T_F = 33$ Mev. Hence, with $S = 8$ Mev we get $m/m^* = 2.1$, which is near enough to the value $m/2$ commonly used for the effective mass.

Evidently the assumption (9) of a quadratic dependence of V_i on p_i is arbitrary. All we can conclude is the fact that the average of V_i is larger than V_F. The main point is the recognition that a decrease of the well depth with increasing momentum of the particle follows directly from two facts: One is the independent-particle motion, the other is the approximate equality of P and S, which is a consequence of the saturation of nuclear forces.

It is not hard to understand in a qualitative way why a momentum-dependent well depth may arise from a nuclear force which has saturation effects. A simple example is the case of repulsive forces at close distances; if a particle moves slowly, it will follow a path avoiding near collisions. When it goes fast, it will penetrate somewhat into the region of repulsion, thus increasing the average potential energy. Another example is an exchange force in which antisymmetric pairs lead to repulsion. The nearest distance of approach between antisymmetric pairs is of the order of the reciprocal relative momentum. Hence, the repulsion will be more effective for a fast-moving particle. This latter example was already treated in 1935 by VanVleck.[5] At that time the independent particle model was in fashion and he has shown that the exchange terms in the average potential energy go to zero with increasing energy of the particle. This gives rise to a momentum dependence of the potential which is quadratic only for momenta small compared to p_F. It reaches a constant value for $p \gg p_F$.

The question arises as to whether the momentum dependence of the well depth can be found directly from experiments. The most direct way would be a study of the nuclear energy levels. The excitation energies are altogether higher if the well depth decreases with increasing momentum. In fact, they should be twice as high as calculated with the normal mass when m^*/m is about $\tfrac{1}{2}$. Unfortunately, however, nuclear spectroscopy is not as direct a way as it seems to find

[4] Note must be taken here of the fact that the nucleus decreases its size by a factor $(A-1/A)^{1/3}$ when one particle is removed. In order to see clearly the effects of this, let us divide the separation of a particle in two steps. First we remove the particle without changing the wave functions of any of the other particles. This requires the energy (7). We then have a nucleus with $A-1$ particles, but with slightly too large a radius. As a second step we compress this nucleus to its normal density. This will raise all kinetic energies slightly, but it also will decrease the potential energy because of the increased density. These two effects must cancel each other almost completely since we know that the actual nuclear density ρ_0 is a stable equilibrium value and hence $\partial E/\partial\rho = 0$ for $\rho = \rho_0$. Hence, (7) is the actual separation energy apart from very small corrections. As to details of the corrections, see R. A. Berg and L. Wilets, Phys. Rev. **101**, 201 (1956).

[5] J. H. VanVleck, Phys. Rev. **48**, 367 (1935).

this effect since one rarely observes the actual energy differences between shells. What one measures are mostly energy differences within one shell, which depend on the finer details of the well shape and are therefore not useful for our problem. The studies of Ross, Mark, and Lawson[6] have shown that the essential features of the low-lying nuclear spectra are not very sensitive to the effective mass.

Levels of higher excitations are difficult to observe and to identify. Once the excitation energy becomes comparable to the distance between shells, the number of levels is so great that a comparison with theoretical predictions becomes hardly feasible. In this connection it is worth mentioning Wilkinson's[7] idea that the giant resonance in photonuclear process might be explained by one-particle transitions, but only if the effective mass is taken into account.

Recently Rand[8] investigated this suggestion quantitatively and has calculated the cross section for nuclear excitation by gamma rays on the basis of an independent particle model in which the particles move in a square well and have an effective mass of one-half. The dependence of these cross sections upon the energy of the gamma rays (he considered only electric dipole transitions) shows the characteristic maxima, usually called "giant resonances." They are almost at the correct energies, although perhaps 10 to 20% lower than the experimental maxima. Without the reduced effective mass, however, the independent particle model gives absorption maxima at energies less than half of the experimental value.[9]

It has been often suggested that the effective mass enters into the contribution of the orbital motion to the magnetic moment. One might suspect that the orbital g factor $(e/2mc)$ is changed by replacing m by m^*. The experimental evidence does *not* bear this out. The slope of the so-called Schmidt lines directly gives the orbital g value, and it seems that the experimental magnetic moments reproduce the slope fairly well. It is true that the values do not lie very close to the lines because of configuration interaction, but a change of the g factor to twice its value seems out of the question. The problem of the effect of the momentum dependent potential on the orbital g factor is difficult and has not been completely cleared up. It is certainly incorrect, however, merely to change m into m^* in the expression for g. One must keep in mind that there is *no* Hamiltonian in which the m^* appears. The mass m^* only enters into the calculation of the expectation value of the energy in the independent particle model state Ψ. The exact Hamiltonian H contains the mass m and not m^*, and the corresponding operator of the orbital magnetic moment is $(e/2mc)L_p$, where L_p is the orbital angular momentum of the protons and m is the actual mass. Therefore, if Ψ is a good wave function, one should use the actual mass in calculating the magnetic moment, and the original Schmidt lines are correct. However, Ψ might be a good wave function for energy computations, but a bad one for magnetic properties.

Recently Blin-Stoyle[10] has suggested the possibility that the effective mass might play a role in determining the moment of inertia of a deformed nucleus. According to Inglis[11] this moment is given by

$$J=\sum_k \frac{|(0|L_x|k)|^2}{E_k-E_0},$$

where L is the operator of the angular momentum around the axis of rotation, and the indices 0 and k denote the ground state and an excited state of the deformed nonrotating nucleus. E_0 and E_k are the respective energies. If this formula is applied to a system of independent particles of mass m moving in an ellipsoidal potential well, one gets for J the so-called "rigid" moment of inertia, which is the one corresponding to the rigid rotation of the mass distribution. This moment is larger than the observed ones. Blin-Stoyle points out that an increase of the excitation energies of the system (and that is just what $m^*/m<1$ means in effect) would decrease the resulting value of J. Figure 1 shows that the observed values of nuclear moments of inertia seem to reach just one-half of the rigid moment for large deformations. This is just the value expected for $m^*/m=\frac{1}{2}$. If this argument is correct, it is an example showing that interactions between particles can reduce the moment of inertia. Bohr and Mottelson have mentioned that a similar effect takes place when one introduces explictly attractive interactions between particles outside closed shells. They have shown that



FIG. 1. The ratio of the observed values of the moments of inertia to the value for rigid rotation are plotted against the deformation parameter β. Values taken from Bohr and Mottelson, Kgl. Danske Vid. Medd. 30, No. 1 (1955).

[6] Ross, Mark, and Lawson, Phys. Rev. 102, 1613 (1956) and a subsequent paper to be published soon.
[7] D. Wilkinson, Proceedings of the Amsterdam Conference on Nuclear Reactions, 1956, to be published in *Physica*. *Proceedings of the Glasgow Conference on Nuclear Physics* (Pergamon Press, London, 1955), p. 161.
[8] S. Rand, Phys. Rev. (to be published).
[9] J. L. Burkhardt, Phys. Rev. 91, 420 (1953).
[10] R. Blin-Stoyle, Nuclear Phys. 2, 169 (1956).
[11] D. R. Inglis, Phys. Rev. 96, 1059 (1954).

the small values of J for small deformations are very probably caused by such effects.

III

We now turn to the discussion of nuclear reactions. The recognition of the relatively independent motion of nucleons in nuclear matter has had its effect upon the interpretation of nuclear reactions. Under the impact of the successes of the shell model it seems questionable to assume that a nucleon, which enters the nucleus from outside, will share its energy immediately with all other constituents as assumed previously. The old description of a nuclear reaction as proceeding in two stages, the formation of a compound nucleus and the subsequent independent decay, is an idealization which cannot be considered to be valid in all cases.

We would like to propose a more general scheme for the description of nuclear reactions that allows for phenomena which do not fall within the framework of the old two-stage description; Fig. 2 presents a graphic representation of such a scheme. We divide the nuclear reaction in three successive stages, the independent-particle stage (I.P. stage), the compound-system stage (C.S. stage), and the final stage. In the first stage we find the incident particle interacting with the target nucleus, but in this stage the nucleus acts upon the particle as a potential well. The particle enters this well without losing its distinct individuality; it is refracted, and partially reflected, at the surface because of the change of potential. In order to provide for the subsequent events in the next stages, the potential must have a real part, V_1, and also an imaginary part, V_2. Then the first stage is described by an incident wave which is not only scattered by the nucleus, but also partially absorbed. "Absorption" in this picture means that the particle disappears from the entrance channel such that it can no longer be considered as existing as an independent particle distinct from the target nucleus. It is just this absorption which leads to the actual nuclear reaction. The scattering which takes place in this stage is called "shape-elastic" scattering. Both scattering and absorption in this stage are given directly by the scattering and absorption of a wave in a complex potential V_1+iV_2. Hence, the events in the

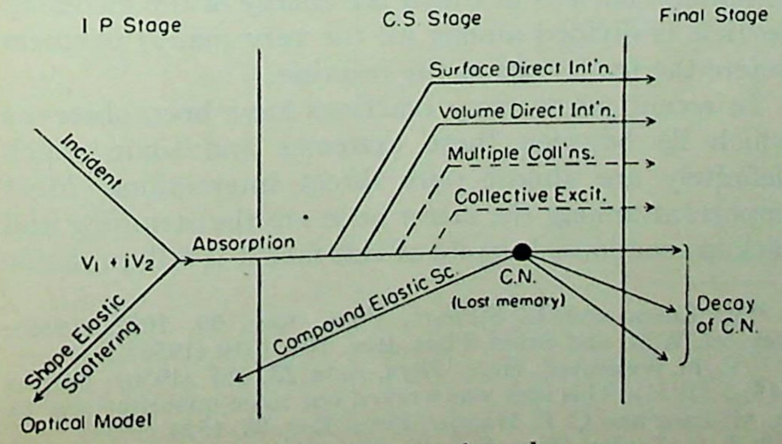


FIG. 2. Nuclear reaction scheme.

first stage are described by the optical model of the nucleus. It is a description of what happens in the entrance channel only.

The second stage contains the events which cause the absorption in the first stage. This "absorption" comprises any effect in which the particle leaves the entrance channel and therefore has undergone an interaction with the target nucleus which cannot be described by a potential only. The state of the system after the particle has been removed from the entrance channel will be called a compound system (C.S.). It describes the situation which exists when the particle interacts with the target to a more intimate extent than can be described by a potential in the entrance channel. This can happen in many ways. The incident particle can collide with a nucleon in the target (direct interactions); it can set up some collective motion as surface vibrations or nuclear rotations. The concept of C.S. is somewhat more general than the concept of "compound-nucleus" (C.N.) as used before in the two-stage Bohr description of a nuclear reaction. The C.N. is characterized by the fact that the state of the incident particle is indistinguishable from the state of any other nucleon; it has completely "coalesced" with the target. The C.S., however, is a state in which some energy exchange between target and projectile has taken place, regardless of the role of the incident particle. Hence the C.S. includes the C.N., but contains also other forms of interaction, namely, all those in which the incident particle is removed from the entrance channel.

There remains a question of what part of the interaction between the incoming particle and the target nucleus can be described by a potential in the entrance channel and what part leads to an absorption in respect to this channel and therefore to a C.S. This problem is not yet clearly understood. After all, the phenomenon described by a potential well is also an interaction, but the target acts in this case only as a whole. Although it might be partially excited when the particle is within the nucleus, it remains in its original state after the particle has left.

The third and final stage contains the processes in which the reaction products separate from each other. In some respects it is similar to the first stage, since the emitted particle can be considered as an outgoing wave from a potential well representing the residual nucleus.

The first stage is the one we know most about. As long as one restricts oneself to the first stage only, all details of the nuclear reaction are buried in what appears as an absorption from the entrance channel. Hence an understanding of this stage provides only a very rough picture of the reaction. Nevertheless, it gives information about the total cross sections, about the elastic scattering, although only about the shape-elastic part, and about the reaction cross sections. Another limitation comes from the short duration of this stage. The time the particle spends within the nucleus before being "absorbed" is only of the order of a few

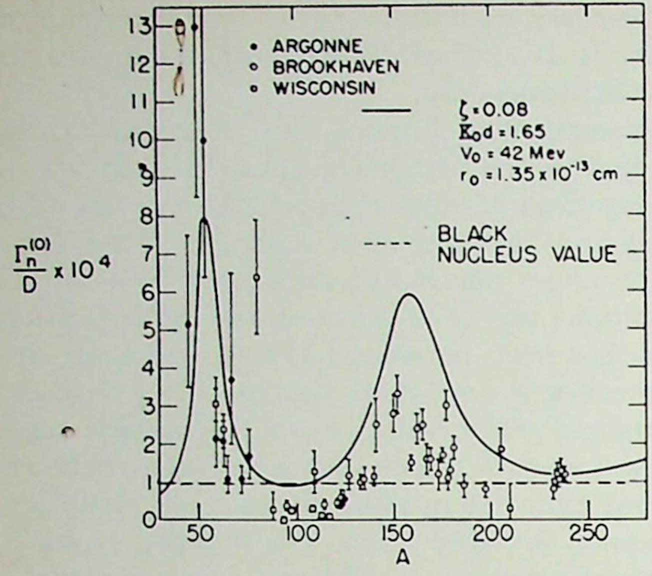


FIG. 3. The average value of Γ_n/D (neutron width) divided by level distance as a function of A. The theoretical curve is calculated with a potential as given by (12) with $R=r_0A^{\frac{1}{3}}$ and $K_0=(2mV_0/\hbar^2)^{\frac{1}{2}}$.

traversals of the nuclear well or even less. It then is impossible within the first stage to ascribe an energy to the particle which is better defined than by the margin given by the uncertainty principle, and this margin is about 0.1 Mev or more. Hence the optical model can only give information which is an average over a relatively large energy region; it therefore cannot account for the closely-spaced resonances which are observed in so many nuclear reactions.

Within these limitations the optical model description of the first stage has been very successful. It can reproduce the main features of the total neutron cross section as a function of energy and nuclear radius by means of a simple potential square well with rounded edges. It can predict reaction cross sections; for example, the theory predicted before the experiment that the average neutron width Γ_n divided by the level distance D for low energies, when plotted against the nuclear radius, should have maxima at certain radii. These are the ones at which a standing wave can be set up within the nucleus $[R=(n+\frac{1}{2})\Lambda$, where Λ is the wavelength inside, and n is an integer]. Figure 3 shows that the experiments have borne out this prediction. It also can predict the size and the angular dependence of the elastic scattering of neutrons and protons.[12]

Actually all one can calculate by this method is the shape elastic scattering which differs from the observed one by the compound elastic scattering. However, this difficulty is not important for protons, in which case the compound elastic part is negligible. In the case of neutrons, the compound elastic cross section, which is important only at lower energies, can be estimated from the difference between the observed reaction cross section and the calculated absorption cross section. The angular dependence of the compound scattering is nearly isotropic. It is gratifying that the difference between the observed elastic scattering and the calculated shape-elastic one can be explained completely by the experimentally determined compound-elastic part.[13]

The best potential for the description of the first stage of a nuclear reaction is at present

$$V=V_1+iV_2, \quad V_2=\zeta V_1$$
$$V_1=-V_0(1+\exp[(r-R)/d])^{-1}$$
$$V_0=43\pm 3 \text{ Mev}, \quad d=0.5\pm 0.1\times 10^{-13}\text{ cm} \qquad (12)$$
$$R=(1.27A^{\frac{1}{3}}+0.6)0.10^{-13}.$$

The value of ζ which determines the imaginary part depends strongly on the energy. For low energies up to a few Mev one finds $\zeta\approx 0.08$. At higher energies it increases sharply and reaches about 0.2 around 10 Mev and perhaps 0.4 to 0.5 at still higher energies. These values and the sharp rise with energy can be qualitatively understood on the basis of the independent particle model and the Pauli principle. The latter prevents the exchange of energy and momentum between the incoming particle and another nucleon if the resulting states are occupied by other particles. With increasing energy of the incoming particle, this will happen less frequently, and therefore more opportunities exist for the incoming particle to interact.[14]

Although ζ is the most energy-dependent variable in the potential (12), the other values are not completely energy independent either. In particular, the well depth V_0 has a tendency of decreasing with increasing energy. Some proton measurements indicate that $V_0\sim 36$ Mev at an incident energy of 30 Mev, and an analysis by Taylor[15] has shown that it reaches a value of about 15 Mev at very high energies. This is just what one would expect from previous considerations as to the momentum dependence of the well depth.

The second stage of nuclear reactions is much less well understood. Here we face a varied range of phenomena, which can be grouped between the two following extremes: One is the direct interaction in which the incoming particles hit one single nucleon in the nucleus, and one of the two partners leaves the nucleus without interfering at all with the other nucleons; the other extreme is a formation of a real compound nucleus in which the energy of the incoming particle is divided among all (or very many) partners before the final stage of the reaction.

In recent years many reactions have been observed which lie between these extremes and some which definitely are almost pure direct interactions. Most important among the latter ones are the stripping and pickup reactions. In spite of this fact it is still probable

[12] R. D. Woods and D. S. Saxon, Phys. Rev. **95**, 577 (1954); **101**, 506 (1956); Fujimoto and Hussain, Phil. Mag. **46**, 542 (1955); Burge, Fujimoto, and Hussain, Phil. Mag. **1**, 19 (1956).

[13] L. Rosen and L. Stewart, Phys. Rev. **99**, 1052 (1955); Beyster, Walt, and Salmi, Phys. Rev. **104**, 1319 (1956).

[14] V. F. Weisskopf, Helv. Phys. Acta **23**, 187 (1950); Science **113**, 1 (1951). This idea was worked out more quantitatively by A. M. Lane and C. F. Wandel, Phys. Rev. **98**, 1524 (1955).

[15] T. B. Taylor, Phys. Rev. **92**, 831 (1953).

that compound nucleus formation plays an important part in most reactions. There is good experimental evidence[16] that the low-energy neutrons emerging from nuclear reactions have a "Maxwellian" distribution in energy (if the energy is high enough to allow for many excited states of the residual nucleus) which is isotropic in space. Hence the statistical description of the second stage as a "heated" compound nucleus is a good approximation. One can conclude from these observations that, for incident energies up to 25 Mev (neutrons or protons), a C.N. is formed with at least 80 to 90% probability.

In recent years attention was mostly concentrated upon those reactions which do not go via C.N. Especially in the reactions in which charged particles are emitted, the C.N. necessarily plays a minor role, since the potential barrier discriminates against charged particle emission after the C.N. is formed. Most of the emissions of charged particles must come from the nuclear surface, because it is a region in which the barrier is somewhat lower. Figure 4 shows a plot by N. Rosen, comparing the observed proton spectrum in a (n-p) reaction on Zn with the one calculated from the statistical theory. The maximum is much lower than the barrier at the nuclear radius (6 to 7 Mev) would admit. Hence, most of these protons must have been produced in an outer surface region.

This example should serve to show that the problems

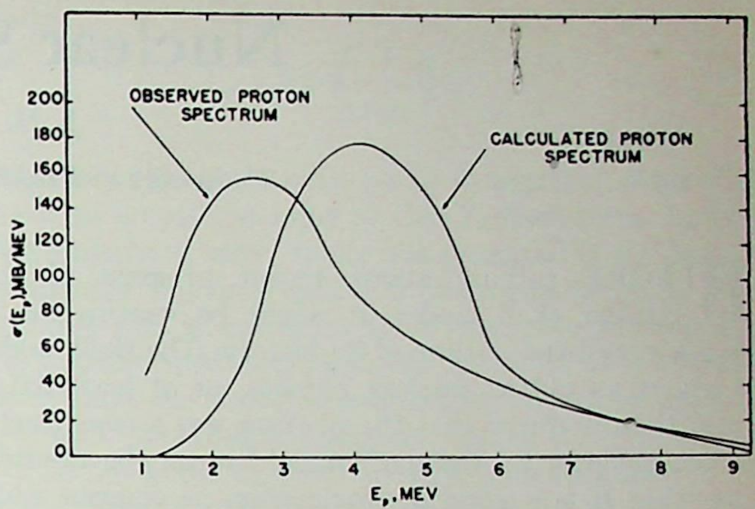


FIG. 4. Comparison of calculated and observed proton spectra from $Zn^{64}+n+14.1$ Mev.

of nuclear reactions are far from being understood. Many more similar examples could be quoted. The situation is rendered difficult because of the likelihood that the properties of the nuclear surface enter essentially into those reactions which do not go via C.N. What happens in the surface is even less understood than what happens in the interior of the nucleus.

The increasing complexity of experimental facts from nuclear reactions shows how far we are yet from a systematic understanding of what is going on in a nucleus during a reaction. It is to be expected that the picture will be clearer when some of the fundamental questions of the structure of nuclear matter are answered.

[16] E. R. Graves and N. Rosen, Phys. Rev. **94**, 224 (1953); N. Rosen and Stewart, Phys. Rev. **97**, 224 (1955); E. R. Graves and R. W. Davis, Phys. Rev. **97**, 1205 (1955).

Nuclear Shell Models

J. H. D. Jensen

University of Heidelberg, Heidelberg, Germany

BEFORE talking about recent progress of the nuclear shell model, it might be worthwhile to give a very brief outline of its history. The shell model is about as old as nuclear physics, or at least dates from the discovery that the neutron was a constituent of the nucleus. By the shell model I mean the assumption that it is a good approximation to describe each nucleon as moving in an average field of force produced by all the other nucleons. This system is so analogous to the Bohr atomic model that it is not necessary to go into details. One obtains a series of degenerate levels which are to be filled according to Pauli's principle. Whenever one encounters two levels which are widely separated, the filling of the lower one means something like a shell closure.

Much has been said about the average potential in Dr. Weisskopf's talk. For the moment, let us assume that it is spherically symmetric; this assumption is not as well founded as in the atomic case. In Dr. Mottelson's talk it will be seen what a great variety of models is accessible when this simplifying assumption is discarded. The other assumption was (at least in the early stages of the shell model, that is, in the early 30's) that the average potential should be a function of the coordinates only. We now know that it might be more reasonable to include a velocity dependence of the potential, say with a term which is proportional to the square of the momentum, but it should still be scalar. This new idea about the best average potential does not affect the general arguments seriously.

I should now like to say a few words about the rather great success which the old shell model attained through the efforts of many theoreticians, especially Wigner and his co-workers, in explaining many facts about the light nuclei which were generally known with rather poor accuracy at that time. I mention as an example, that the shell closures of He^4 and O^{16}, and to a certain extent of Ca^{40}, were the first theoretical predictions about the particular stability of certain nuclei. The shell model provided a very neat explanation of the surprising fact that until far up in the atomic table there was only one element (argon) which had no odd mass isotope: One knew that the only odd mass argon isotope which might be expected to be stable was $_{18}A_{21}{}^{39}$, this would lie on the slope of increasing neutron excess with increasing charge, but the shell model explained rather quantitatively the instability of this isotope just by the fact that the twenty-first neutron would have to be brought into a new shell and would be very loosely bound. Hence the isobar $_{19}K_{20}{}^{39}$ has a lower binding energy in spite of the greater Coulomb energy, and therefore there is no odd mass isotope of argon.

These are only a few examples for the general success of the early shell model. Yet the model somehow got out of vogue. This was partly for the reason that there were other nucleon numbers which were distinguished in the same way as the numbers 2 in helium, 8 in oxygen, and 20 in calcium, from the natural radioactive decay one knew that there were similar phenomena at the numbers $Z=82$ and $N=126$, and from the table of stable nuclei together with the abundance chart it could be concluded that $Z=50$, and $N=50$ and also $N=82$ played a similar role. However, these numbers never arose as shell closures from a reasonable assumption about the average potential even if one included velocity dependence. The main reason, however, was the tremendous success of Niels Bohr's compound nucleus model in explaining the data on nuclear reactions. The underlying idea of the compound nucleus is that all nucleons somehow equally share the responsibility for the properties of the nuclear states involved, and Bohr rather convincingly suggested that this were a necessary consequence of the close packing and the short range of the nuclear forces. It was generally believed that this feature not only held for the excited levels in nuclear reactions, but also for the lowest levels, and I think that mainly in consequence of this generalization the shell model was not considered seriously for about ten years.

This morining we learned about Dr. Brueckner's confidence that these arguments may be overcome, and that the independent particle model can be justified in spite of close packing, and the short range forces, and even for forces with a repulsive core. The principal objections may therefore no longer be serious, and I hope that in the near future this point can be completely cleared up. As long as it is not convincingly shown that the Hartree model is a good approach even for these extreme forms of nuclear interactions, the success of the shell model is like explaining magic by miracles.

At any rate, some years ago Dr. Goeppert-Mayer and, independently, my friends Haxel and Suess in Germany dug up the old shell model. At that time we were not very hopeful that we could learn very much about elementary nuclear interactions in this way, but we thought that even a poor model might be useful in finding some correlations and orientations in the rapidly increasing wealth of isolated nuclear data. We thought the situation might be rather similar to that in organic chemistry before quantum mechanics was developed. At that time, the experimental chemists

could not wait until the physicists had developed an entirely consistent and fundamental theory of matter, and they had to use concepts of their own, like valence or the Kekulé model. We investigated the shell model in this spirit and we were quite surprised that it appears to be possible to obtain some fundamental physical information from it.

The first reason why we reconsidered the shell model was the rapidly increasing evidence for the particular behavior of the nucleon numbers mentioned above; since they were not explained at that time, the name "magic numbers" was used in the literature. Beyond that, I should like to mention two other aspects. First of all there was the different behavior of the magnetic moments of odd proton and odd neutron nuclei, namely that nuclear moments of odd neutron nuclei did not show any dependence on the angular momentum (they are by no means greater for large angular momenta than for small angular momenta) whereas there is definitely an increase of the odd proton moments with increasing angular momentum. This suggests that only the orbital momentum of the odd charged particle, contributes to the magnetic moment of odd mass nuclei. I will not go into details, but in addition there is a grouping which even suggested a difference in orientation of the spin of the odd particle with respect to its orbital momentum. Thirdly, there was another point which Niels Bohr emphasized in my discussions with him, namely the absence of rotational levels. At that time none were known, and there are still no rotational levels with moments of inertia of the whole nucleus. Now in a model in which each particle is moving in an average field, there is no place for a rotation of the nucleus as a whole because any angular momentum must be given by the orbital momenta of the single particles. Of course, there might be superpositions, but the shell model was the simplest way to get rid of the surprising fact that there were no rigid nuclear rotations.

We next tried to find the proper "magic numbers" and, as is well known, they could be obtained uniquely by one additional assumption about the spherically symmetric potential. To the purely radial potential one has to add a strong spin-orbit coupling of the form

$$f(r)\boldsymbol{\sigma}\cdot\mathbf{r}\times\mathbf{p}.$$

This additional strong spin-orbit coupling gives the right ordering of levels.

At that time, the existence of a potential of this type was an arbitrary assumption, although such a term does arise as a first order relativistic effect if one considers the motion of a spin $\frac{1}{2}$ nucleon to be described by a Dirac-like equation with a spherically symmetric scalar field $V(r)$. In the Pauli approximation this leads to a term of the type

$$\left(1-A(\boldsymbol{\sigma}\cdot\mathbf{L})\frac{1}{r}\frac{d}{dr}\right)V(r),$$

where

$$A=\left(\frac{h}{2mc}\right)^2.$$

The sign of the term with $(\boldsymbol{\sigma}\cdot\mathbf{L})$ is negative, otherwise the term is just the same as the Thomas term. In the electromagnetic case, where the potential is the fourth component of a four-vector, it has a positive sign. The negative sign is just what is necessary to account for the experimental data, but the factor A is too small by at least one order of magnitude, and therefore the spin-orbit potential was originally an *ad hoc* hypothesis.

I do not know to what extent the Brueckner theory is now able to give a spin-orbit coupling of this type from conventional nuclear forces as they are suggested by meson theories (for example, from the tensor forces). At the moment one can hardly give a simple reason for the occurrence of this peculiar potential. Of course, the spin-orbit splitting need not necessarily be described by such a term, but it is certainly the simplest one.

I should like to emphasize the experimental evidence for this spin-orbit coupling from the scattering of neutrons or protons by helium. Helium can be considered as a very spherically symmetric nucleus, it has no excited states up to its dissociation energy of about 20 Mev, and therefore I think that if the shell model should work at all, it would be most usefully applied to the interaction of the helium nucleus with an incoming proton or neutron. The analysis of the proton scattering showed very different behavior of the phase shifts for the $p_{\frac{1}{2}}$ and $p_{\frac{3}{2}}$ states. There is a rather sharp resonance which means the $p_{\frac{3}{2}}$ phase shift goes through $\pi/2$ at about 2.5 Mev (laboratory system) whereas the $p_{\frac{1}{2}}$ phase shift shows a very gradual slope. This obviously indicates a very strong spin-orbit interaction. These phase shifts were confirmed at Minnesota by the experimental observation of the polarization of proton beams, through double scattering in the classical arrangement by which Barkla showed polarization of x-rays.

The first scattering on helium produces the polarization and the next scattering provides the analysis of the polarization. The amount of polarization predicted by the phase shifts (as obtained from the angular dependence of the single scattering at different energies) was quantitatively confirmed.

This experiment has been extended, and the whole phase shift analysis confirmed by experiments at Illinois. Jentschke and co-workers let the first scattering occur at rather high energy (5.5 Mev) and before the second scattering the protons are slowed down to 2.5 Mev (this does not affect the polarization because the slowing down is caused by ionization). Because of the different effects of the phase shifts at the different energies, the theory predicts a reversal of the intensities in two symmetric directions, as compared with the Minnesota experiment. Complete agreement with the

theoretical prediction was found. I think that the spin-orbit interaction for a nucleon moving in nuclear matter is now an established fact.

Breit and his co-workers tried to obtain these phase shifts through the whole region from zero to 15 Mev under the assumption that the proton or neutron moves in a velocity independent, spherically symmetrical potential with a spin-orbit term of the Thomas type superimposed. For a Gaussian well one had to take the spin-orbit coupling 30 times the Thomas value, and for a square well potential, one had to take even 50 times the Thomas value. The phase shifts were described over the whole energy interval with velocity independent central forces. Perhaps this is an accident but we cannot draw conclusions until these calculations are repeated with velocity dependent potentials.

I shall not discuss how these shell model predictions were useful in giving the spins and parities, etc., not only of stable nuclei, but also of radioactive nuclei, in the classification of β decays, in the explanation of the phenomenon of isomerism, etc. But I will discuss some recent more detailed calculations which were made with the aim of getting more quantitative agreement with experimental facts, and of getting information about elementary nuclear forces.

The closed shell nucleus is very trivial, and equally simple are nuclei with one neutron or one proton missing from a closed shell, but as soon as there are several particles outside closed shells the situation is very peculiar. If the shell model were really a very good approximation the particles in an incompletely filled shell could couple their spins in various ways, even if one takes the requirements of all the symmetry principles: isotopic spin, and Pauli's exclusion principle into account. The prediction would be that very close to the ground state there should be other levels with different spins, so that such nuclei should have a very confusing spectrum. From the fact that this is not the case, one sees that simple shell model functions are a crude approximation and that there are certainly correlations between the nucleons. The adopted procedure is quite similar to what was done in the theory of atomic shells and the early nuclear shell model: One takes the shell model functions as first approximation, describes the interactions by pairs of particle forces, and then one calculates what mixtures of different shell model configurations gives the lowest energy. One hoped that at least the bulk of the eigenfunction could be represented by only one configuration. That is actually not true, and to solve this problem there have been carried out a great deal of numerical calculations by very many theoreticians.

It is impossible to talk about all the details which have been obtained. However, I should like to mention two which I happen by accident to know best and which I think are rather consistent. One is the work by Kurath on the whole $p_{3/2}$ and $p_{1/2}$ shell, that is for all nuclei between helium and oxygen. In pure jj coupling first the $p_{3/2}$ shell would be filled, and after that one had to fill the $p_{1/2}$ shell. But all other forces besides the spin-orbit force (which we have not really understood as yet) tend to produce a mixture of these configurations so that in the simplest case the Li^6 ground state should be a linear combination of $(p_{3/2})^2$, $(p_{3/2},p_{1/2})$, and $(p_{1/2})^2$.

Calculations show that, with reasonable forces between pairs of particles and a single particle spin-orbit coupling superimposed, one arrives at a very considerable mixing of configurations. The situation is intermediate between a pure LS configuration and a pure jj configuration. Kurath showed that with rather good over-all agreement not only the ground states but also the excited states can be obtained.

If one allowed for admixtures of further configurations which arise from higher single particle states, the agreement could certainly be improved, because this procedure is a legitimate and systematic way of taking all correlations between the particle positions into account. But it is reassuring that with p configurations alone the experimental data can be reproduced so well. To give this agreement the single particle spin-orbit force has to increase systematically from the beginning of the shell (He^5) towards the end of the shell (N^{15}). Such a behavior of the phenomenological $f(r)\boldsymbol{\sigma}\cdot\mathbf{L}$-force follows from the assumption that it arises from spin-orbit interactions between pairs of particles. The beginning of the shell (He^6) is therefore better represented by a pure LS configuration, whereas the end of the shell (N^{14}) resembles more closely a pure jj configuration.

Not only the energies and quantum numbers of the nuclear states are well accounted for, but also the matrix elements for radiative and β transitions agree fairly well with the observed data. This is a more sensitive test for the quality of the eigenfunctions, because it is well known that the energies (which can be obtained from a minimum principle) usually come out quite well even with poor eigenfunctions, whereas the transition matrix elements are much more sensitive. The long lifetime of C^{14}, i.e., the small matrix element for its β transition to N^{14}, which was a worrying puzzle to the theoreticians for many years, is now well understood. The matrix element becomes accidentally so small through a destructive interference between the contributions of the various configuration which represents the ground states of C^{14}, respectively N^{14}. From central forces between pairs of particles and $\boldsymbol{\sigma}\cdot\mathbf{L}$ forces alone one cannot obtain a configuration mixture with such amplitudes and phases that the matrix-element is smaller by a factor 10^{-3} than the value calculated from a simple shell model configuration. However, it was shown by the Harwell group that the addition of a small tensor interaction between the two particles, of sign and magnitude as in the deuteron, is sufficient to yield precisely such a configuration mixture

that the calculated β-transition matrix element vanishes, whereas the ordering and spacing of the excited levels is practically not disturbed.

Kurath's p-shell results agree well with those obtained by the Harwell group, and this group achieved equally satisfactory results for the nuclei with one to three nucleons beyond the O^{16} shell. In this case they find that an appreciable mixing of the $d_{\frac{5}{2}}$, $s_{\frac{1}{2}}$, and $d_{\frac{3}{2}}$ configurations results from the same type of internucleon forces. Again good agreement with observed data not only of the level schemes but also of electromagnetic moments and of the transition probabilities is obtained without additional f-state configurations. All these calculations were carried through with shell model eigenfunctions obtained from a velocity independent potential and a $f(r)\boldsymbol{\sigma}\cdot\mathbf{L}$ term.

In these calculations one difficulty in the concept of the shell model has also been eliminated, namely, the additional degrees of freedom of the "center" around which the auxiliary single particle potential is assumed to be rotationally invariant, which has no physical significance. The total nuclear wave function should be invariant under translation, whereas the determinant of the A single particle wave functions gives an oscillation of the center of gravity around the nonphysical "center." For the nuclear ground states this unrealistic feature can be removed. However, if one considers all possible states of all shell-model configurations one would obtain states which have no physical meaning. The simplest example is an "excited state" of He^4. The ground state is an $\{s^4\}$ configuration. For excited states a configuration $\{s^3p^1\}$ could be considered. Among the properly antisymmetrized functions there appears one which can be written as $\{s^2p^1\}_{J=1}=\mathbf{R}\cdot\{s^4\}$, where $\mathbf{R}$ is the center of gravity vector of the four nucleons. This would describe an unphysical state. Only after elimination of all such spurious shell model states does one obtain level schemes which agree with the observed ones.

I could give only a brief summary of a few of the numerous calculations which have shown that the shell model can account for a great number of quantitative nuclear data as well as for general features of nuclear structure if correlation effects are properly accounted for by the introduction of mixed configurations. This practical success must however be solidified by a general proof that the Hartree method is self-consistent for the nucleus.

Collective Motion in the Nucleus

Ben R. Mottelson

CERN *Theoretical Study Division at the Institute for Theoretical Physics, University of Copenhagen, Copenhagen, Denmark*

For many purposes it is possible to consider each nucleon as moving, almost independently, in an average nuclear field generated by all the other nucleons. The existence of such independent particle motion in the nucleus gives rise to a nuclear shell structure which is similar in many respects to the familiar atomic shell structure. There is, however, at least one very important difference between the field in which the electrons move and that in the nucleus. In the atom the field is dominated by the attraction of the heavy central nucleus. The fact that the nuclear field is generated entirely by the nucleons themselves implies that the nucleus will be much less stable against oscillations in shape. Already in the early considerations about nuclear structure, it was pointed out by Bohr and Kalckar[1] that this relative instability in shape should imply the existence of low-frequency shape oscillations as an important mode in the nuclear dynamics. The understanding of most of the low energy nuclear properties requires an analysis of the interplay between the collective shape oscillations and the independent particle degrees of freedom in the nucleus.[2]

In general, the period for the shape oscillations of lowest frequency is considerably longer than the period for independent particle motion in the nucleus. One may thus employ an adiabatic approximation in solving the equations of motion. That is, we can proceed in two steps: first we solve for the intrinsic motion subject to a constraint which specifies a particular shape and orientation. In practice we may attempt to satisfy this constraint approximately by solving for the motion f nucleons in a potential which possesses the re- uired shape and orientation. The eigenvalues, $E_i(\alpha)$, f this first part of the problem are functions of the parameters, α, which define the nuclear shape. These functions are called the potential energy surfaces. They play a role very similar to the potential energy surfaces in molecules. Next, we consider the additional energy which the system acquires when we let α vary slowly with time. Expanding the energy in a power series in $\dot{\alpha}$ we obtain the collective Hamiltonian

$$H=E_i(\alpha)+\tfrac{1}{2}B_i(\alpha)\dot{\alpha}^2. \qquad (1)$$

The neglect of the nondiagonal couplings proportional to $\dot{\alpha}$ is justified only if there are no degeneracies or near-degeneracies in the spectrum of the intrinsic motion. In general it is only the even-even nuclei that are free from such degeneracies. I shall mainly confine myself in this talk to these even-even nuclei for which the above especially simple collective Hamiltonian is adequate. Unfortunately, I shall not have time, within the scope of this talk, to discuss the many very interesting effects which arise in the odd-A nuclei.

In the early discussions of the nuclear collective properties, an attempt was made to estimate the functions $E(\alpha)$ and $B(\alpha)$ by employing a liquid drop model. Thus, $E(\alpha)$ would depend on the nuclear "surface tension" which in turn could be related to the observed nuclear binding energies, and B might be calculated from the kinetic energy associated with surface oscillations of an irrotational fluid.

We now know, however, that the functions $E(\alpha)$ and $B(\alpha)$are influenced in an essential manner by the nuclear shell structure. Let us consider, for example, the function $E_0(\alpha)$ which represents the lowest potential energy surface in an even-even nucleus. Beginning with a closed shell nucleus, we know that the system possesses an especially great binding energy which is associated with the degeneracies of the nucleon orbits in a spherical potential. Any attempt to distort the system from the spherical shape is very costly in energy, and the potential energy curve for this configuration is thus stable at the spherical shape and rises very steeply as we go away from this shape (curve a, Fig. 1).

Adding additional nucleons beyond the closed shell, the potential energy curve will be a result of the competition between the particles outside closed shells, which exert a polarizing effect on the nuclear shape, and the particles in closed shells which prefer the spherical shape. The existence of this important polarizing effect of the individual nucleons outside of closed shells was first recognized by Rainwater.[3] He pointed out that since the orbital motion of a single nucleon is mainly confined to a plane perpendicular to the direction of its angular momentum vector, the particle will exert a centrifugal force in this plane tending to distort the nucleus.

The polarizing tendency of the particles outside of closed shells is modified in an important way by the residual forces which are not included in the average nuclear field. These tend to couple the nucleons into a state of definite total angular momentum J. Mayer,

[1] N. Bohr and F. Kalckar, Mat.Fys.Medd.Dan.Vid.Selsk. 14, no. 10 (1937).

[2] A. Bohr, Mat.Fys.Medd.Dan.Vid.Selsk. 26, No. 14 (1952). D. L. Hill and J. A. Wheeler, Phys. Rev. 89, 1102 (1953). A. Bohr and B. R. Mottelson, Mat.Fys.Medd.Dan.Vid.Selsk. 27, No. 16 (1953). The present talk is largely based on Chapter V of the review article by Alder, Bohr, Huus, Mottelson, and Winther, Revs. Modern Phys. 28, 432 (1956). This last reference also contains a more complete bibliography and compilation of relevant experimental data.

[3] J. Rainwater, Phys. Rev. 79, 432 (1950).

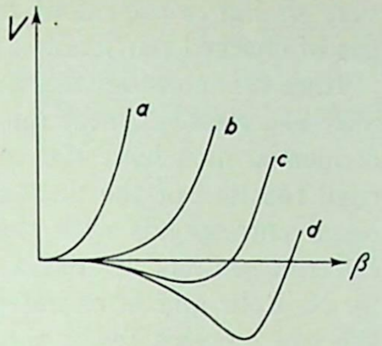


FIG. 1. Potential energy curves for the lowest configurations in even-even nuclei.[2]

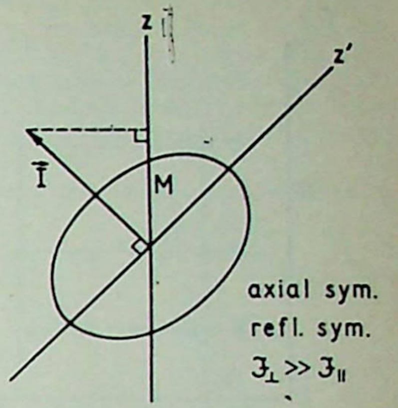


FIG. 2. Coupling scheme for the lowest rotational band in an even-even nucleus.

Flowers, Racah, and others have shown that the short range attractive forces, which act in the nucleus, favor a state of $J=0$ for an even-even nucleus. Thus, in a spherical potential ($\alpha=0$), the particles in an even-even nucleus will have $J=0$, a spherical density distribution, and no net polarizing effect on the nuclear shape. However, when we consider intrinsic motion in a nonspherical field, J is no longer a constant of the motion, the particles become decoupled from each other and exert their polarizing effect.

Thus if we add a few particles beyond a closed shell we obtain a potential energy curve which remains stable at the spherical shape, as a consequence of the residual interactions, but which is much softer against distortion away from the spherical shape as a consequence of the polarizing tendencies of the added nucleons (curve b, Fig. 1). Continuing to add still more nucleons beyond the closed shell, the spherical shape continues to become less and less stable, until, if there are enough nucleons outside of closed shells, the spherical shape may become unstable and the lowest energy of the nucleus is then associated with a nonspherical shape (curve c, Fig. 1). With still more nucleons added, the magnitude of the equilibrium deformation increases and the stability of the system with respect to displacements from the deformed equilibrium also increases (curve d, Fig. 1.).

I would like now to discuss the characteristic collective excitation spectra which are associated with these different types of potential energy surfaces.

The situation is most simple for nuclei with nonspherical equilibrium shapes, so I shall start by discussing the spectra of the nuclei in the regions farthest from closed shells. The regions in which nonspherical nuclei have been observed are roughly defined by $A>24$, $150<A<190$, and $A>220$; these are just the regions farthest from the closed shell configurations. Such systems, possessing a nonspherical shape, exhibit a very low energy collective motion which simply corresponds to the reorientation of the nucleus in space with preservation of the shape and intrinsic structure. Indeed the nuclei in the above regions do exhibit rotational spectra which for the lowest intrinsic state in the even-even nuclei have the form

$$E_{\mathrm{rot}}=\frac{\hbar^2}{2\mathfrak{J}}I(I+1) \quad I=0,2,4,6,\cdots \text{even parity.} \quad (2)$$

The nature of this rotational spectrum tells us quite a bit about the symmetry properties of the nuclear shape. Thus the dependence of the nuclear energy on $I(I+1)$ implies that we have to do with the rotations of a symmetrical top; the fact that we most often observe only the even spin and even parity states in these spectra[4] implies that usually the nuclear shape is symmetric with respect to reflection in a plane perpendicular to the symmetry axis and passing through the center of the nucleus; and finally, the absence of other states in the rotational band which would correspond to rotation about the symmetry axis implies that the effective moment of inertia for such rotations is at least an order of magnitude smaller than that for rotations about an axis perpendicular to the nuclear symmetry axis. The nuclear coupling scheme is sketched in Fig. 2. These symmetry properties of the nuclear shape are related to the nuclear potential energy surfaces. It has indeed been shown that the nuclear shell structure will almost always prefer such axially symmetric shapes. The smallness of $\mathfrak{J}_{\parallel}$ is a special case of a more general result which we shall discuss further a little later.

The experimental evidence on the rotational spectra of the even-even nuclei is shown in Fig. 3. This gives the ratio of the measured excitation energies of the excited states to the energy observed for the first excited, 2+, states in these nuclei. According to the above expression we would expect the second excited state to have $I=4+$ and an energy $3\frac{1}{3}$ times that of the 2+ state, the third and fourth states should have $I=6+$ and 8+, and energy respectively, 7 and 12 times the 2+ energy. The experimental data agree quite well with this interpretation; in the regions of the largest deformations the energy ratio of the 4+ to the 2+ states agree with the above expression within a few tenths of a percent. The deviations are always negative, increase with increasing I, and with the approach to the transition regions where the deformed shape is no longer a stable equilibrium. Such deviations can be interpreted in terms of centrifugal distortion, which slightly changes the nuclear shape as a function of the rotational frequency.

[4] In some cases, especially around $A=224$, odd-spin odd-parity rotational bands have been observed. It has been suggested by R. Christy that these be interpreted in terms of a deviation of the nuclear shape from reflection symmetry.

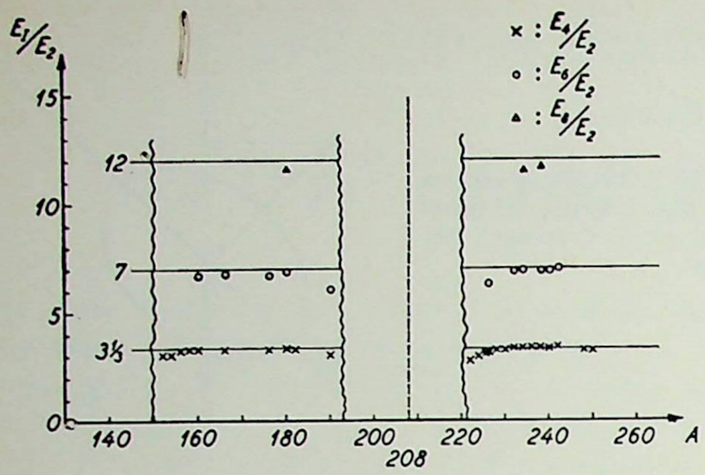


FIG. 3. Experimentally observed energy ratios for excited states in rotational bands in even-even nuclei.[2]

Besides the simple energy expression governing the rotational spectrum, there are also intensity rules that govern the relative strength of β or γ transitions to different members of a rotational band. These intensity rules follow directly from the geometric nature of the rotation, in the same way that one obtains intensity rules for the relative strength of the transitions in a fine structure or a hyperfine structure multiplet. The experimental evidence supports these intensity rules within the experimental accuracy, which is about 10% in the best studied cases.

While the relative energies and intensities in a rotational band are geometrical quantities which are independent of the detailed nature of the intrinsic motion, the effective moment of inertia, $\mathfrak{J}$, appearing in the rotational energy, is a dynamical quantity which depends in an essential way on the intrinsic structure of the system. We may attempt to estimate $\mathfrak{J}$ by employing a method first suggested in this connection by Inglis.[5] He pointed out that, since $\mathfrak{J}$ represents the additional kinetic energy which the nucleons must have in order to follow the rotation of the nucleus, we can estimate $\mathfrak{J}$ by considering the motion of particles in a rotating nuclear field. Due to the equivalence of a rotation to an external magnetic field, this problem is very similar to the calculation of the diamagnetism of a gas of charged particles.

If we first consider this problem in the limit where we may use classical mechanics we find that the moment of inertia must have the value $\mathfrak{J}_{\text{rig}}$ corresponding to a rigid rotation of the nuclear density distribution. This result corresponds with the statement that a classical electron gas has no diamagnetism, as was first shown by N. Bohr and later independently by van Leeuwen. We can see this result most easily by looking at the system in the rotating coordinate system in which the potential remains constant in time, but where we must add the appropriate Coriolis forces to the Hamiltonian. However, the Coriolis forces do not change the isotropy of the velocity distribution. There is thus no net current in the rotating coordinate system; in the space fixed system the whole nucleus is rotating as a rigid body. This result is independent of any details of the intrinsic structure such as the mean free path of the particles in the potential.

When we consider the actual nucleus we find that there is a very important quantum-mechanical correction to the above result, which is a consequence of the tendency of the intrinsic structure to go into a state of $J=0$ as the nuclear deformation, α, goes to zero. As mentioned previously, this tendency is a consequence of the residual forces in the nucleus. Since a state with $J=0$ has a spherical density distribution, its energy is independent of the orientation of the nuclear field, and thus it does not contribute to the moment of inertia. Even for the appreciable eccentricities encountered in the most deformed nuclei, this single quantum state with $J=0$ constitutes a significant part of the intrinsic state of motion and thus the nuclear moments of inertia are appreciable below the value $\mathfrak{J}_{\text{rig}}$ and become smaller as the eccentricity becomes smaller.[6] The experimentally measured moments of inertia are shown in Fig. 4 in units of the moment $\mathfrak{J}_{\text{rig}}$. The abscissa is the nuclear eccentricity, β, deduced from the nuclear quadrupole moment determinations. The solid curve is obtained from an extremely simplified model which provides an interpolation between the small and great deformation limits. This interpolation expression depends on the strength of the residual interactions as measured by the parameter v.

As we approach the closed shell configurations the deformations get smaller, the moments of inertia get smaller, and the accuracy of the simple rotational energy spectrum gets poorer. Finally, we reach a point where the nuclear potential energy surface no longer has a nonspherical equilibrium and then the nuclear coupling scheme changes in an essential manner. The nucleus now has a spherical equilibrium shape and the system cannot exhibit simple rotational motion; the low-frequency collective degrees of freedom now correspond to

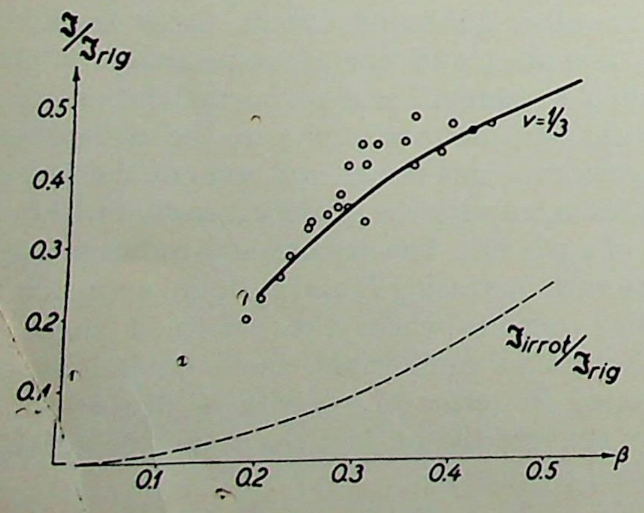


FIG. 4. The magnitudes of the observed moments of inertia[6] are plotted as a function of the deformation parameter β.

[5] D. R. Inglis, Phys. Rev. 96, 1059 (1954).

[6] For a more detailed discussion of the nuclear moments of inertia, see A. Bohr and B. R. Mottelson, Mat.Fys.Medd.Dan. Vid.Selsk. 30, No. 1 (1955).

TABLE I. Data on nuclear vibrations around a spherical equilibrium shape. The table[2] lists all cases in which the second excited state is known to have spin and parity 2+ (the first excited states are all also known to have 2+ character). The second, third, and fourth columns list, respectively, the energies of the first excited state, the second excited state, and the ratio of these two energies. The fifth column gives the ratio of the observed cross section for Coulomb excitation to that expected for a single proton transition. The sixth column gives the ratio of $M1$ to $E2$ radiation in the cascade transition between the second excited state and the first. The last column gives the ratio of the reduced transition probabilities for $E2$ radiation for the crossover decay to ground and the cascade decay of the second excited state.

Nucleus	E_2 (Mev)	E'_2 (Mev)	E'_2/E_2	$\frac{B(E2;0\to2)}{B_{sp}(E2)}$	$(M1/E2)_{2'\to2}$	$\frac{B(E2;2'\to0)}{B(E2;2'\to2)}$
$_{26}Fe^{58}$	0.81	1.62	2.00		0.2	0.01
$_{28}Ni^{60}$	1.33	2.18	1.64	17		(3×10^{-3})
$_{30}Zn^{64}$	1.00	2.27	2.27	15		(0.1)
Zn^{66}	1.05	2.40	2.29	11		(0.05)
$_{34}Se^{76}$	0.55	1.19	2.17	44	~1	0.1
$_{36}Kr^{82}$	0.77	1.45	1.88			(0.01)
Kr^{84}	0.9	1.9	2.1			>0.1
$_{40}Zr^{92}$	0.93	1.83	1.97			(0.05)
$_{44}Ru^{100}$	0.54	1.36	2.52	22		(0.05)
Ru^{102}	0.47	1.10	2.34	45		(0.15)
$_{52}Te^{122}$	0.57	1.26	2.21	26	0.1	0.01
Te^{126}	0.65	1.40	2.16	17		(0.004)
$_{54}Xe^{126}$	0.39	0.86	2.20			(0.01)
Xe^{128}	0.46	0.99	2.15			(0.01)
$_{78}Pt^{192}$	0.32	0.61	1.90		0.025	0.004
Pt^{194}	0.33	0.62	1.88	50	small	0.01
Pt^{196}	0.35	0.69	1.97	38	0.05	$<4\times10^{-4}$
$_{80}Hg^{198}$	0.41	1.09	2.66	29	0.7	0.04
$_{84}Po^{214}$	0.61	1.38	2.26	13	>2	

quadrupole vibrations around the spherical equilibrium shape. Each quantum of vibration carries two units of angular momentum, since we are dealing with quadrupole oscillations. If we assume small amplitude oscillation, the motion will be approximately harmonic and the excitation spectrum of an even-even nucleus would have the form indicated in Fig. 5. Of course, higher order terms in the collective Hamiltonian will modify the exact equality of the energy spacings and remove the degeneracies.

As was first pointed out by Scharff-Goldhaber and Weneser,[7] the low-lying excitation spectra of the even-even nuclei, outside of the above regions where rotations are observed and excluding the few nuclei immediately adjacent to closed shells, agree very well with this picture. Thus the first excited states all have spin 2 and even parity. The energy of this state varies in a regular manner with A and in the sense suggested by the qualitative considerations about the potential energy curves. The frequency decreases regularly as we go away from closed shells, corresponding to the decreasing stability of the spherical shape caused by the polarizing effect of the nucleons outside of closed shells.

Coulomb excitation of these first excited states has been observed in about fifty nuclei, and in every case the cross section has been appreciably greater than would be expected for the excitation of a single proton; the enhancement ranges between about a factor of ten to fifty. This provides very direct evidence that we are dealing here with an excitation mode that involves the motion of an appreciable number of nucleons. The Coulomb excitation cross sections also increase as we go away from closed shells, reflecting with the expected increase in the amplitude of the oscillations of shape.

The second excited vibrational states have been observed in many cases and their energy is found to vary between 2 and 2.5 times that of the first excited state (with two or three unusual exceptions lying outside this interval). The observed second excited states have in all cases $I=0+$, $2+$, or $4+$. The whole triplet is not usually observed because the states are populated in β or γ decays which obey very strict selection rules on the allowed spin changes.

The vibrational character of the second excited state is especially shown by the γ decay of the second excited $2+$ states. These states may decay by $M1$ or $E2$ radiation to the first excited $2+$ state or by $E2$ radiation to the $0+$ ground state. If all matrix elements were of the order of single particle values, the $M1$ cascade decay would be the strongest by factors of the order of 100. However, the observed $M1$ cascade radiation is usually no stronger than the $E2$ cascade and is sometimes appreciably weaker. This smallness of the $M1$ transition matrix elements follows at once from the fact that the fundamental excitation mode involved is a quad-

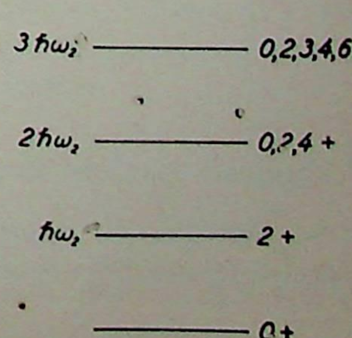


FIG. 5. Excitation spectrum corresponding to quadrupole oscillation of small amplitude. Energies are shown on the left, and spins and parities on the right.

[7] G. Scharff-Goldhaber and J. Weneser, Phys. Rev. 98, 212 (1955).

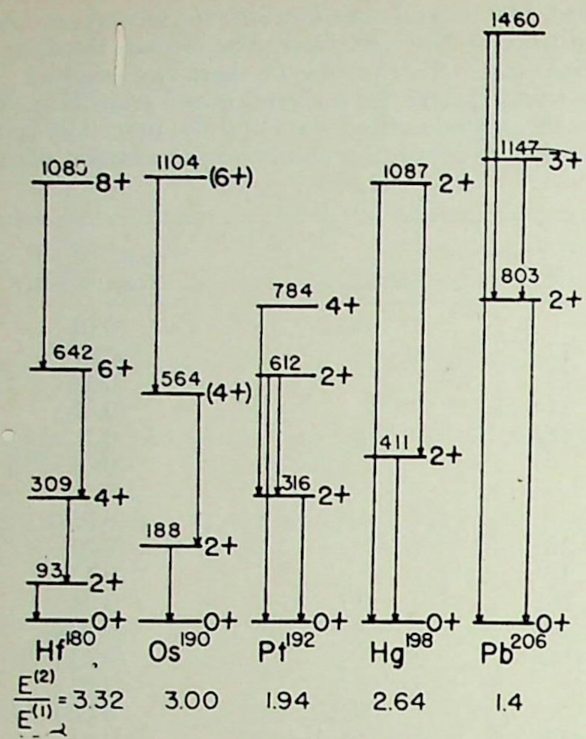


FIG. 6. Representative low-energy excitation spectra of even-even nuclei in the region $170 < A < 208$. The ratio of the energy of the second excited state, $E^{(2)}$, to that of the first excited state, $E^{(1)}$, is shown under each level scheme. In the region of nonspherical nuclei the limiting expression (2) predicts that this ratio should be $3\frac{1}{3}$; in the region where the spherical shape is stable, the assumption of approximately harmonic vibration implies that this ratio should be close to 2 (see Fig. 5).

rupole motion, even though $\Delta I = 0$, and thus in this description the $M1$ matrix element vanishes.

The reduced $E2$ transition probabilities for the cascade and crossover decay of the second excited 2+ states are observed to differ by an appreciable factor. The reduced transition probability for the cascade decay is observed to range from 10 to 1000 times greater than that for the crossover transition. The existence of this selection rule also follows as a simple consequence of the above description of these states. Since the electric quadrupole transition operator is linear in the quadrupole shape parameter, α, it cannot change the number of vibrational quanta by more than one.

The empirical evidence on the vibrational spectra of even-even nuclei is summarized in Table I. This contains all the cases in which the second excited states are known to have spin and parity 2+.

To summarize, the general picture of the lowest excited states of the even-even nuclei is as follows:

In the regions farthest from closed shells the nuclei possess a nonspherical shape as a consequence of the polarizing effect of the nucleons outside of closed shells. The lowest excitations are then rotations with the spectrum (2). As we approach towards a closed shell the magnitude of the nuclear eccentricity decreases and the moment of inertia also decreases. At a point which may be still quite distant from the closed shell configuration, the spherical shape becomes a stable equilibrium as a consequence of the residual forces between the nucleons and the whole nuclear coupling scheme changes. The lowest excitations now correspond to approximately harmonic shape oscillation of quadrupole type. With the further approach to a closed shell configuration the frequency of such shape oscillations increases. Finally, in the immediate neighborhood of a closed shell configuration the collective vibrational frequencies may become comparable to those of independent particle motion. It is then no longer possible to employ an adiabatic approximation as in the derivation of (1). In such cases it is most appropriate to treat completely all the degrees of freedom of the particles outside of the closed shells, as has been done by Inglis, Elliott, and Flowers, Redlich, Ford, Levinson, Pryce, and others.

The general sequence of spectra is illustrated in Fig. 6. The levels of Hf^{180} follow the rotational spectrum quite accurately, as is the case for all the nuclei in this region (Fig. 3). With Os^{190} deviations from the rotational spectrum (2) are about ten percent. The transition to vibrational spectra occurs at Pt^{192}. All the observed spectra from this point to the immediate neighborhood of the closed shell configuration of Pb^{208} correspond to approximately harmonic vibrations.

Finally, by the time we come to Pb^{206} the collective frequencies have become higher than those of individual particle motion. The observed low-lying excited states of this nucleus have been shown[8] to correspond well with the two-neutron excitations expected from the known single particle levels available in this region.

[8] M. H. L. Pryce, Proc. Phys. Soc. (London) A65, 773 (1952); D. E. Alburger and M. H. L. Pryce, Phys. Rev. 95, 1482 (1954).

Nuclear Reactions

A. M. LANE
Atomic Energy Research Establishment, Harwell, England

1. INTRODUCTION

THE richness of phenomena encountered in the study of nuclear reactions is revealed by the colorful terminology in current use: cloudy crystal ball model, knock-out processes, boil-off processes, stripping and pickup reactions, and so on. In a brief survey with any pretensions to coherence, it is impossible to make detailed mention of all phenomena. Consequently, the present review will be purposely restricted to the central features of nuclear reactions. Coulomb excitation, heavy ion reactions, fission, and other such specialized subjects will not be discussed. Furthermore, little will be said about the older and well-known subjects such as the Breit-Wigner formula and the statistical theory of the compound nucleus. Rather we will stress the developments of recent years, with particular emphasis on the cloudy crystal ball (or complex potential) model and direct reactions.

Ideally, the task of a comprehensive theory of nuclear reactions is to predict the cross sections for all the individual energetically-allowed processes that can be initiated by the bombardment of a given target nucleus with a given projectile. In the search for such a theory, it is convenient to begin by limiting oneself to a more modest goal, namely the prediction of the cross sections for scattering and absorption.

When the target nucleus and projectile come together in a bombardment, there are two basic types of events that can occur as a result of their interaction. Either the nuclei are scattered by each other without change in relative energy or internal structures (scattering event), or else they exchange energy and change their internal structures (absorption event).* In the latter case, the event implies the formation of a "compound system" of all the particles.[1] This may decay quickly (direct process) or after a long time interval (compound nucleus process). Evidently, any given individual cross section is determined by the relative probability for the decay of the compound system into the appropriate final nuclei. A theory giving all these relative probabilities can be called a theory of "individual cross sections" as opposed to one which only gives only the two basic cross sections for scattering and absorption. Section 2 is devoted to discussion of the cross sections for diffraction and absorption. This is followed, in Sec. 3, by a review of the present theory of individual cross sections. Finally, in Sec. 4, some remarks are made on the general contemporary situation in the study of nuclear reactions.

2. CROSS SECTIONS FOR SCATTERING AND ABSORPTION

The sum of the scattering and absorption cross sections defined in the Introduction must evidently constitute what is normally understood to be the "total" cross section. However, it is not true[2] that the two cross sections separately can be exactly identified with what one normally calls the "elastic" and "nonelastic" (or "reaction") cross sections. The reason for this comes from the possibility that, of the totality of absorption events which constitute the absorption cross section, some of them may ultimately lead to the production of the original bombarding pair of nuclei. In the scheme which divides the total cross section into elastic and nonelastic, such events would be included in the former, not the latter. These special events are said to account for the "compound-elastic" part of the elastic cross section. This distinguishes them from the "scattering" or "shape-elastic" events which constitute the rest of the elastic cross section. Fortunately, except at energies of less than a few Mev, the compound-elastic cross section is quite negligible. Thus in the following, we will identify the usual quantitative definitions of elastic and nonelastic cross sections with the shape-elastic and absorption cross sections, respectively. Later on [Sec. 2(d)] we will describe the special situation arising at low energies.

At this stage we make an important specialization in the discussion, namely it will be assumed that the projectile in the bombardment is a nucleon. Clearly, a necessary preliminary to the basic understanding of reactions initiated by deuterons, alphas and other composite particles is the understanding of reactions initiated by simple single particles.

2(a) Parametric Analysis of Cross Sections

Let us, for simplicity, ignore the presence of the intrinsic nucleon spin. The shape-elastic and absorption

* Strictly speaking, there may also be the third possibility of a charge in relative energy without a change in internal structure. This will be so if energy can be stored in collective motion which preserves internal structure. For instance, if the target nucleus has a nonspherical shape, it may be set into simple rotational motion. We ignore this possibility until we make special mention of it in Sec. 4(b).

[1] V. F. Weisskopf, Address at Amsterdam Conference on Nuclear Reactions (July, 1956).

[2] Feshbach, Porter, and Weisskopf, Phys. Rev. 96, 448 (1954).

cross sections then have the well-known forms

$$\sigma_{\text{sh.el}}(\theta)d\Omega=\sigma_{\text{cl}}(\theta)d\Omega=\frac{\pi}{k^2}\left|\sum_l (2l+1)(1-S_l)\frac{P_l(\cos\theta)}{(4\pi)^{\frac{1}{2}}}\right|^2 d\Omega$$

$$\sigma_{\text{abs}}=\sigma_{\text{nonel}}=\frac{\pi}{k^2}\sum_l (2l+1)(1-|S_l|^2),$$

where k is the wave number of the relative motion of the colliding nuclei, l labels the orbital angular momenta of relative motion, and S_l is the scattering amplitude for the lth partial wave defined by the following asymptotic form of the wave function of relative motion at large distances of separation

$$\frac{1}{r}\{e^{-i[kr-(l\pi/2)]}-S_l e^{+i[kr-(l\pi/2)]}\}P_l(\cos\theta).$$

When the nucleon spin is taken into account, these formulas become somewhat more complicated. For instance, for a zero spin target nucleus, each S_l is replaced by two scattering amplitudes $S_{l+\frac{1}{2}}$ and $S_{l-\frac{1}{2}}$ for two possible values, $J=l\pm\frac{1}{2}$. In addition, there is a new cross section to be considered, the polarization cross section, which depends on the differences $(S_{l+\frac{1}{2}}-S_{l-\frac{1}{2}})$. The presence of nucleon spin is only significant when there is some force coupling the spin to the orbital motion, so that $S_{l+\frac{1}{2}}$ and $S_{l-\frac{1}{2}}$ are not equal. Experimentally, it appears that this coupling is not very large, although it is certainly necessary to take it into account if the observed differential cross sections are to be fitted in better than a qualitative manner. (Also, evidently, it is necessary to consider it if one wishes to fit the observed polarization cross sections.) For the present discussion, it is convenient to ignore the spin-orbit coupling.

At any given energy the set of complex S_l form a doubly infinite number of parameters that are, in principle, deducible from experiment and which are only restricted by the conditions $|S_l|\leqslant 1$. We might guess that any theory of reactions containing a similar number of parameters is always capable of fitting the data.[3] In situations of physical interest, only a restricted number of l waves are involved in nuclear collisions because of the finite size of nuclei. In a given collision, there are only $\sim 2l_{\max}$ parameters to be considered, where $l_{\max}$ is the l value of the highest partial wave involved in the collision. If a theory contains a similar number of parameters, then it can always be made to fit the data. Such a theory gives no physical insight into the mechanism of reactions. It only gives a "representation" of the data, but does not provide evidence for or against a particular model, although it may be inspired by a model.

[3] J. A. Wheeler, Phys. Rev. 99, 630(A) (1955).

2(b) Complex Potential Representation

Many successful attempts have been made to fit observed cross sections by representing the interaction of incident nucleons with target nuclei by a spherical single particle potential well $-V(r)-iW(r)$, where $V(r)$ is the real, refracting part and $iW(r)$ is the imaginary, absorbing part. For any such potential, one solves the Schrodinger equation for each l wave, thereby determining the amplitudes S_l. In the high energy limit, where many l waves are involved in the collision, the formulas for $\sigma_{\text{el}}(\theta)$ and σ_{abs} naturally take on the forms characteristic of the "optical model" that can be used in this limit.[4] S_l is evaluated as $\exp(ik\int nds)$ where the integral is taken along the optical path corresponding to impact parameter $b=l/k$, and where n is a complex refractive index defined by

$$n^2-1=(V+iW)/E,$$

E being the relative energy of collision. It is usual to write

$$n=1+\frac{1}{k}\left(k_1+\frac{iK}{2}\right),$$

where k_1 and K are real quantities and K is the "absorption coefficient" equal to the reciprocal of the mean free path against collision. If W is small compared to $V+E$, then

$$\frac{k_1}{k}\approx\left(1+\frac{V}{E}\right)^{\frac{1}{2}}-1$$

$$\frac{K}{k}\approx\frac{W}{E\left(1+\frac{V}{E}\right)^{\frac{1}{2}}}.$$

These relations are correct to a sufficient approximation in cases of physical interest.

In order to fit the data with the complex potential model, it is evidently necessary to restrict the potential to approximately the size of the target nucleus if the theoretical value of $l_{\max}$ is to be made to coincide with the experimental one. This still leaves one with freedom to adjust the shapes of $V(r)$ and $W(r)$, and this freedom surely means that, at any given energy, the data can always be fitted by a suitably chosen complex potential.† At very high energies, where the elastic scattering closes up into the forward direction, there is effectively only one experimental quantity, *viz.*, σ_{abs}, and the

[4] Fernbach, Serber, and Taylor, Phys. Rev. 75, 1352 (1949).

† For instance, the quantity $1-|S_l|^2$ giving the absorption in the lth wave is essentially $\sim\int W(r)|\psi_l(r)|^2 d\tau$; i.e., the volume integral of $W(r)$ weighted by the square modulus of the wave function $\psi_l(r)$ of the lth wave. Thus one can give σ_{abs} any desired value by adjusting the value of $W(r)$ near the impact distance for the lth wave where $|\psi_l|^2$ is peaked. However it must be stressed that no *rigorous* proof has been given of the conjecture that the set of S_l can always be fitted by a suitable choice of potential $V(r)+iW(r)$. Such a proof has been reported for only the special case of $W(r)=0$, i.e., no absorption.[3]

corresponding theoretical parameter is an integral over $K(r)$. Thus, at any energy, there seems to be ample freedom in the model to guarantee a fit of the data.

2(c) Evidence for the Optical Properties of Nuclear Matter

The very important question that now arises is: how can one establish whether the complex potential corresponds to physical reality or not? In other words, we want to know if nuclear matter really has optical properties so that the complex potential is a genuine physical model, or whether the complex potential is just to be regarded as no more than a representation of data. Merely to fit the data at a given energy with an adjustable complex potential proves nothing. To confirm that the model has a physical basis, one must go further and compare the experimental values of V and W as functions of radial distance and energy with values computed theoretically using a calculation based on optical model ideas. Generally speaking, such ideas enable one to express V and W in terms of the scattering amplitudes of individual collision processes between the incident nucleon and the nucleons in the target. Since the observed amplitudes for nucleon-nucleon scattering are smooth functions of energy, it follows that a precondition for the physical reality of the complex potential is that V and W are smooth functions of energy. Furthermore, since the radial dependences of V and W are expected to be simply related to that of the nuclear density, the known smoothness of the latter leads to the requirement that V and W be also smooth functions of radial distance.

Experimentally elastic and absorption cross sections have been measured for neutrons and protons incident on many nuclei with bombarding energies up to 1.4 Bev. These cross sections have been analyzed using the complex potential with the following results.

Observed Radial Dependence of V and W

As far as the radial dependences are concerned, all the data are consistent with a shape much like that of the nuclear density distribution with its flat central part and an appreciable surface fall-off distance. These two qualitative features are essential for fitting the data,[5,6] at least in the case of $V(r)$. The radial distribution of W is less well determined and some success has been achieved with a distribution of W concentrated at the nuclear surface.[5] For both V and W there is uncertainty about the details of the shapes deduced from experiment. This arises partly from the neglect of spin-orbit coupling in most attempts to fit the data. To carry out a full analysis of the polarization, differential and absorption cross sections with a complex potential model including a spin-orbit term $U(r)\mathbf{l}\cdot\mathbf{s}$ would be very laborious. Unfortunately it seems that such an analysis must be done if the complex potential model is to be exhaustively tested.

[5] Bjorklund, Fernbach, and Sherman, Phys. Rev. **101**, 1832 (1956).

[6] Melkanoff, Moszkowski, Nodvik, and Saxon, Phys. Rev. **101**, 507 (1956).

Observed Energy Dependences of the Central Depths of V and W

The values of central depths of V and W that are extracted from the data depend on the radii of the radial distributions of V and W. (In principle, such radii should be determinable uniquely from a careful analysis of the data and so V and W could be assigned unique values. In reality, most analyses have not approached this ideal stage.) To a first approximation, the cross sections are determined by the combination $(k+k_1+iK/2)R$, where R is the mean radius and k_1 and K are the "bulk" values of our previously defined quantities. It follows that V and W depend on the choice of R as $(V+E)\sim R^{-2}$, $W\sim R^{-2}$. The values of V and W that we now quote, have been obtained from the reported values by normalizing to the choice $R=1.33A^{\frac{1}{3}}\times10^{-13}$ cm.

For protons,[6] the central depth of V after correction for the mean Coulomb potential is observed to decrease from 55 Mev at zero bombarding energy roughly according to the law $V=55-0.5E$. Above $E=40$ Mev, the values flatten off to $V\sim20$ Mev at $E=100$ Mev and appear to stay at this value up to $E\gtrsim400$ Mev.[7] At higher energies, the values of V are not obtained very precisely from the data.

For neutrons,[2,7] the same remarks apply except that the depth of V at zero bombarding energy[2] is about 10 Mev less than for protons. Such a difference is to be generally expected from the velocity dependence of V. It is easy‡ to see that, if the mean potentials felt by a neutron and a proton are assumed to be the same as functions of *kinetic* energy except for the Coulomb potential V_c, then the two potentials as functions of *total* or *bombarding* energy differ not only by V_c but also by a term which is opposite in sign to V_c and proportional to V_c.

For protons and neutrons the value of the central depth of W increases from $W\sim3$ Mev at zero bombarding energy to $W\sim15$ Mev at $E=32$ Mev.[6] There is a broad maximum with $W\sim20$ Mev centered at $E\sim70$ Mev, followed by a shallow minimum of $W\sim12$ Mev at $E\sim200$ Mev. From then on W increases monotonically passing through $W\sim17$ Mev at $E=350$ Mev[7] and $W\sim70$ Mev at $E\sim1400$ Mev.[8]

[7] T. B. Taylor, Phys. Rev. **92**, 831 (1953); R. Jastrow and R. M. Sternheimer (to be published); F. Mandl and T. H. R. Skyrme, Phil. Mag. **44**, 1028 (1953).

‡ If, for a neutron, $V=-\alpha+\beta T$, where T is the kinetic energy, the corresponding formula for V as a function of *total* energy $E=T+V$ is $V=(-\alpha+\beta E)/(1+\beta)$. For a proton, the formula $V=-\alpha+\beta T+V_c$ leads to $V=(-\alpha+\beta E+V_c)/(1+\beta)=(-\alpha+\beta E)/(1+\beta)+V_c-\beta/(1+\beta)V_c$. This differs from the neutron potential not only in V_c but also in the term $-\beta/(1+\beta)V_c$ which tends to cancel V_c.

[8] R. W. Williams, Phys. Rev. **98**, 1387 (1955).

To appreciate the physical implications of these magnitudes of W, it is convenient to consider the ratio of the corresponding mean free paths $1/K$ to a typical nuclear radius which we will choose to be 7×10^{-13} cm. This ratio exceeds unity only below a few Mev. Above these energies it falls steadily to about 0.25 around $E=40$ Mev, then rises to a maximum of about 0.7 at $E\sim350$ Mev and falls again to 0.3 at 1.4 Bev. Thus the mean free path of a nucleon against collision in nuclear matter is always of the order of the nuclear radius, but only actually exceeds it at low energies.

On the simplest classical particle picture, one expects the mean free path $1/K$ to be simply $(\rho\langle\sigma\rangle)^{-1}$ where ρ is the nucleon density in nuclear matter and $\langle\sigma\rangle$ is a suitable average of the nucleon-nucleon total cross section, taken over the various velocities of neutrons and protons in nuclear matter. This relation follows from the more general one of the classical optical wave model[9]:

$$n-1=\rho k^{-1}\langle 2\pi k_n^{-1}f_n(0)\rangle.$$

Here $f_n(0)$ is the forward part of the nucleon-nucleon scattering amplitude and the preceding expression for K follows on taking the imaginary part of this equation. It is unfortunate that one cannot proceed to calculate k_1 directly from the observed nucleon-nucleon data. The reason is essentially that the singlet and triplet scattering amplitudes are not determined separately by the data as analyzed at present.§ No such difficulty is present in the case of K however and this quantity is straightforwardly computed. The observed values of K above $E\sim100$ Mev are immediately found to be predicted correctly. From the simple formula for K, its energy dependence for $E>100$ Mev ought to be that of the nucleon-nucleon cross sections. The fall-off in the n-p cross section from 100 to 300 Mev and the subsequent rise (due to meson processes) are directly reflected in the observed values of K.

Below $E=100$ Mev the decrease in K can be ttributed to the effects of the Pauli principle which orbids collisions that would otherwise take place. A simple calculation[10] with a degenerate Fermi gas at zero temperature reproduces the observed values rather well if one ignores the velocity dependence of the nucleon potential and takes $V=T_F+8$ Mev, i.e. $E=T-(T_F+8$ Mev) where T is the kinetic energy of the bombarding particle and T_F is the maximum energy in the Fermi distribution. A more consistent calculation which takes the velocity dependence into account reduces the predicted values by nearly an order of magnitude, thereby destroying the agreement with experiment. However the agreement can again be restored by allowing the Fermi gas to have a suitable temperature.[11] The effect of doing this just about cancels the effects of the velocity dependence.

In spite of the absence of an optical calculation of V, the agreement obtained in the case of W between optical theory and experiment seems sufficient to establish the physical reality of the complex potential. The final task is now to explain *why* nuclear matter has apparent optical properties. This is really a problem of nuclear structure rather than nuclear reactions, so this difficult question will not be discussed here. However it seems appropriate to draw attention to the fact that the model's success is somewhat surprising. The model is essentially a classical one, based on the concept of two-particle collisions with energy conservation in each collision. On the other hand, the nucleus is a quantum system of closely packed particles and there are usually particles present in the wave zone for a given collision between two particles, and these would normally lead to nonconservation of energy in the collision. Perhaps the present situation can be optimistically summarized by the observation that the optical model picture is at least internally consistent in the following sense: the model certainly demands as a necessary condition that the localization length (i.e., the wavelength) of the incident particle be much less than the distance between collisions. Assuming the validity of the model, one computes this latter distance (i.e., the mean free path $1/K$) and finds it to be large and to satisfy the necessary condition.

2(d) Special Situation at Low Energies

In the beginning of this section, we mentioned that the absorption and shape-elastic cross sections differ from the usual nonelastic and elastic cross sections by an amount equal to the compound-elastic cross section, thus[2]

$$\sigma_{\rm el}=\sigma_{\rm sh.el}+\sigma_{\rm comp.el}$$

$$\sigma_{\rm nonel}=\sigma_{\rm abs}-\sigma_{\rm comp.el}$$

At energies above a few Mev, $\sigma_{\rm comp.el}$ is reduced to a negligible amount because of the competition of this process with all the many possible decay modes of the compound system. In the case of low energies, when this is not the case, the first thing to notice is that the complex potential model is to be associated with the prediction of $\sigma_{\rm sh.el}$ and $\sigma_{\rm abs}$ and *not* with $\sigma_{\rm el}$ and $\sigma_{\rm nonel}$.‖ (This follows from the physical meaning of the complex potential in which "absorption" signifies an energy-exchanging collision of the incident nucleon with a target nucleon.) Thus, it might be thought that, in order to compare the observed $\sigma_{\rm el}$ and $\sigma_{\rm nonel}$ with the

---

[9] R. Jastrow, Phys. Rev. 82, 261 (1951); N. C. Francis and K. M. Watson, Phys. Rev. 92, 291 (1953).

§ Attempts have been made to compute k_1 by hypothesizing a nucleon-nucleon potential and evaluating the forward scattering amplitude in Born approximation. For example, A. Kind and L. Jess (to be published). See also reference 15.

[10] A. M. Lane and C. F. Wandel, Phys. Rev. 98, 1524 (1955); E. Clementel and C. Villi, Nuovo cimento 10, 176 (1955).

[11] K. A. Brueckner, Phys. Rev. 103, 172 (1956).

‖ The complex potential model makes no specification at all about $\sigma_{\rm comp.el}$, and so does not predict $\sigma_{\rm el}$ and $\sigma_{\rm nonel}$. To determine $\sigma_{\rm comp.el}$, one has to introduce a complete detailed theory of reactions as in Sec. 3 (see reference 13).

complex potential model, one must subtract and add $\sigma_{\text{comp.el}}$ to these quantities. This is correct, but this is not the only modification to be made at low energies. There is a second one, which appears at first sight to be quite different, but really is also due to the non-negligible value of $\sigma_{\text{comp.el}}$. This second modification is that cross sections must always be averaged over the fine-structure resonances before comparing with the complex potential. The complex potential model cannot retain its usual physical meaning when applied to sharply defined energies, if only because the values of V and W must be allowed to fluctuate to follow the resonances. There is a good physical reason[1] why the complex potential cannot be applied to vanishingly small energy intervals, but is restricted to energy intervals including many levels. If the split of the total cross section into $\sigma_{\text{sh.el}}$ and σ_{abs}, as given by the complex potential, is to be a valid physical one, there must be no interference between the compound-elastic scattering and the diffraction-elastic scattering, i.e. the two must be completely incoherent. On a time scale, this means that the incident wave packet must be sufficiently short in time that the shape-scattered wave is well clear of the nucleus before the compound-elastic wave appears at the nuclear surface. The time for the latter is $2\pi\hbar/D$, where D is the mean fine-structure level spacing. Application of the uncertainty relation $\Delta t\Delta E \gtrsim h$ shows that, for no interference, the incident energy must be spread over many resonances.

2(e) "Strength Function" Representation

For some years there has been available a proper general theory of reactions formulated in terms of fine structure resonance levels.[12] This theory is a quantum mechanical theory from the beginning, and so the use of it avoids the awkward questions of interpretation that arise with the use of the complex potential [see Sec. 2(c)]. It is to be expected that this theory would be more formal and forbidding. Nevertheless, it is possible to use the fine structure theory to define absorption and shape-elastic cross sections which are identical in physical meaning to those we have already introduced.[2,13] (The essential point of comparison is that the absorption cross section includes that part of the elastic cross section that is incoherent with the shape-elastic cross section.) These cross sections are given by the formulas of Sec. 2(a) with S_l replaced by its average $\langle S_l\rangle$ taken over many resonances. S_l is related in a simple way to the logarithmic derivative f_l of the wave function[14] evaluated at the nuclear surface, i.e., at the interaction radius $r=a$. It can be shown[13,15] that $\langle S_l\rangle$ is given by the same relation with f_l replaced by a quantity $(R_\infty^{(l)}+i\pi s_R^{(l)})^{-1}$ where $R_\infty^{(l)}$ and $s_R^{(l)}$ are two parameters defined statistically in terms of the widths and energies of the fine structure resonances.¶ s_R is the "strength function," defined as $\langle\gamma_\lambda\rangle^2/D$; i.e., the mean reduced with γ_λ^2 of resonance levels λ, divided by the mean level spacing of these levels. R_∞ is essentially the deviation of the potential scattering phase angle from that due to a hard sphere of the size of the target nucleus of interaction radius $r=a$.

Since the complex potential and the strength function theories both specify the same cross sections, the condition that they give the same cross sections gives relations between the parameters of the two theories. Thus, since we know that the complex potential fits the data with certain values of V and W, we can avoid reanalyzing the data with the strength function theory by using these relations. The condition that the two theories give the same cross sections is equivalent to the condition that they give the same effective value of f. This quantity is $(R_\infty+i\pi s_R)^{-1}$ in the case of the strength function theory. In the case of a square complex potential $V+iW$, with single particle energies E_p (say), f may be written as[13,15]

$$\frac{1}{f}=\sum_p \frac{\hbar^2/Ma^2}{E_p-E-iW}$$

and so, taking the imaginary part

$$s_R=\frac{1}{\pi}\sum_p \frac{(\hbar^2/Ma^2)W}{(E_p-E)^2+W^2}.$$

Thus the form of s_R needed to fit the observed data is a sum of Cauchy terms, each one centered on a single particle energy E_p. It is of considerable interest to try to predict this form of s_R from the theory of nuclear structure. (In particular, the numerical prediction of the width $2W$ of the peaks is especially important.[14]) The starting point for such an attempt is the relation[13]:

$$R_\infty+i\pi s_R=R(E+i\delta),$$

where $R(E)=\sum_\lambda\gamma_\lambda^2/E_\lambda-E$ is the "R function" of the fine structure theory[12] and where δ is an energy much larger than the mean level spacing. The presence of δ removes the singularities in R and makes it smooth. Eventually δ must be allowed to go to zero. Now we may write[15]

$$\gamma_\lambda^2\left(\frac{\hbar^2}{Ma^2}\right)^{-1}=(\sum_p \langle\lambda|p0\rangle)^2\approx\sum_p |\langle\lambda|p0\rangle|^2$$

[12] E. P. Wigner and L. Eisenbud, Phys. Rev. 72, 29 (1947); P. L. Kapur and R. E. Peierls, Proc. Roy. Soc. (London) A166, 277 (1937); G. Breit, Phys. Rev. 69, 472 (1946).

[13] E. P. Wigner, Proc. Cambridge Phil. Soc. 47, 790 (1951); Ann. Math. 53, 36 (1951); 55, 7 (1952); R. G. Thomas, Phys. Rev. 97, 224 (1955).

[14] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley and Sons, Inc., New York, 1952).

[15] Lane, Thomas, and Wigner, Phys. Rev. 98, 693 (1955).

¶ Since the fine-structure theory is general, there must be as many parameters in it as there are in the set of S_l. For each (complex) S_l, there correspond the two parameters $R_\infty^{(l)}$ and $S_R^{(l)}$. From this point in the discussion, we drop the label l.

where $\langle\lambda|$ is the resonance state λ and $|p0\rangle$ is the product state of the single particle state p and the ground state $|0\rangle$ of the target nucleus. The last inequality follows on assuming random signs for the amplitudes $\langle\lambda|p0\rangle$. Thus

$$R(E+i\delta)=\left(\frac{\hbar^2}{Ma^2}\right)\sum_p\sum_\lambda\frac{\langle\lambda|p0\rangle^2}{E_\lambda-E-i\delta}$$

$$=\left(\frac{\hbar^2}{Ma^2}\right)\sum_p\left\langle p0\left|\frac{1}{H-E-i\delta}\right|p0\right\rangle.$$

The matrix elements can now be developed in a perturbation series[16] in terms of $H-H_0$ where H_0 is the Hamiltonian for the state $\langle p0|$. Each term in the series is a sum over processes in which the incident particle is scattered through several states before returning to its original state p. One can argue that, if the matrix elements of $H-H_0$ connect states separated by energies up to some energy $\mathcal{E}$, then under reasonable conditions on $\mathcal{E}$, one process contributes more than all the others.[16] This is the process in which the incident particle is returned to its original state by each alternate scattering. The contributions of such processes to all terms in the perturbation expansion can be summed in closed form to give just the required Cauchy form. In this form,

$$W(E)=\pi\frac{\text{Av}|\langle p0|(H-H_0)|p't\rangle|^2}{D(E)},$$

where $|p't\rangle$ labels those product states with energies near E, and the average is taken over these states. $D(E)$ is the mean level spacing at energy E.

This calculation seems to be a more precise derivation of the same result obtained from more intuitive considerations[17] in which one calculates the energy shift in the state $|p0\rangle$ due to the interaction $(H-H_0)$. The previous approximation in which the incident particle is returned to its original state by each alternate scattering corresponds to the Tamm-Dancoff method of radiation theory in which single photons are allowed to be created and destroyed. The energy shift from this theory contains an imaginary part just like the W given by the above formula. Although quantitative numbers are unreliable, the qualitative values of W from this formula are in agreement with experiment.[17] In particular, the Pauli principle is decisive in reducing the value of W at low energies.

The strength function theory does not depend in any way on the complex potential model and the calculation of cross sections just described may be said to give a direct description of these cross sections without any reference to the complex potential. However the latter model is a very simple physical model that is known to fit the data. Therefore a better description of the strength function theory would perhaps be that it gives a proper quantum mechanical basis for the complex potential model. In this connection we note that the foregoing expression for $W(E)$ can be reduced to the form

$$W(E)=\frac{\hbar^2}{2M}\rho\langle\sigma_{\text{Born}}\rangle(k+k_1)$$

which is very similar to that implied by the optical theory in Sec. 2(c). The only difference between the two expressions is that, where the optical formula contains the actual nucleon-nucleon total cross sections, the other contains Born approximation expressions for these quantities. It is not clear at present whether this difference is a result of approximations made in the use of the strength function theory (the selection of leading terms) or whether it implies that the classical optical model ought to be modified to the extent of this difference. At high energies the Born approximation reproduces the observed total cross sections and the two expressions become the same.

3. INDIVIDUAL CROSS SECTIONS

So far, this review has only mentioned the prediction of the cross sections for shape-elastic scattering and absorption of incident particles. Now the attempts to decompose the absorption cross section into its component parts (the individual cross sections) will be described.

3(a) Statistical Theories of Reactions

In general, a given bombardment may result in the production of very many different products. In the course of producing these, a very large number of nucleon-nucleon collisions occur and one must resort to some type of statistical treatment. Until a few years ago, it was assumed that, for all but high energies (say > 50 Mev), the incident nucleon is absorbed into the target nucleus as soon as it crosses the surface and forms a long-lived compound state. The qualitative features of this state have been described so many times[14] that there is no need to repeat them here. The prediction of any individual cross section on this model has the form of a product of (i.e. the geometrical cross section πa^2 at sufficiently high energies) and the probability of decay into the particular final products specified by the cross section. This probability is computed on a statistical basis assuming that the intrinsic probabilities for decomposition into all individual final products are the same. One prediction of this theory is that the spectrum of emitted particles of a given type (say protons) should be essentially Maxwellian ($e^{-E/T}$) except for a low energy cutoff due to barrier penetration.

An essential condition for the validity of this "strong absorption" picture is that the mean free path of the incident nucleon against collision be much smaller than

[16] C. Bloch, Nuclear Phys. (to be published).
[17] M. Cini and S. Fubini, Nuovo cimento 10, 75 (1955).

the dimensions of the target nucleus. As we have already seen, the use of the complex potential model in analyzing observed cross sections shows that this condition is *not* satisfied. The empirical mean free paths obtained in this way are of the order of the nuclear radius with the consequence that the incident particle, after its first collision, still has an appreciable probability of escaping from the nucleus. It was fortunate that this information from the complex potential model appeared when it did, because there was growing experimental evidence against certain aspects of the "strong absorption" picture,[14] and it was not known at what point the theory was at fault. It is fairly straightforward to modify the "strong absorption" model to allow for the long mean free paths. Using a classical treatment of particle-particle collisions, one imagines[18] the incident particle to collide with a target nucleon, then the two nucleons collide with other nucleons and so on, so that a cascade develops. After each collision there is a certain small probability that a nucleon will escape directly. For a reasonably high energy, say 50 Mev, there are a very large number of collisions, so that it is almost certain that one or two nucleons will escape directly. The end of what might be called the "cascade stage" of the reaction comes when all individual nucleons have less than the requisite energy for escaping out of their mean potential well. At this point we can imagine that a true compound nucleus is formed and that this lives for a relatively long time before eventually decaying.** ††

The method of calculation for the "cascade stage" of the reaction depends on what one wishes to predict. For instance, it is clear that the products of the first collision will have a strong memory of the incident direction and energy. The products of second collisions will have less memory i.e. they tend to appear at wider angles and lower energies. With each collision, memory is progressively lost so that particles from the sixth or seventh collisions emerge nearly isotropically with low energies. If one observes the high-energy particles in the forward direction, it follows that these are mostly due to first collisions. It is quite easy to make an analytical calculation[19] of the energy and angular spectra of these nucleons (using the "impulse approximation"). This can be extended to the second collision,[19] but the method becomes tedious for higher order collisions. This means, in particular, that it is not useful for predicting the excitation energy left in the compound nucleus. For following these higher order collisions, one usually takes resort to the type of mathematics known as Monte Carlo calculations, in which one follows many cascades through the nucleus and builds up statistical results.[18] Such calculations have been done for an incident nucleon energy of 100 Mev on a nucleus of mass 64, under the assumption that a nucleon cannot escape after its energy falls below the maximum Fermi gas kinetic energy in the target nucleus plus 8 Mev.[20] Out of 200 particles, 30 pass through the nucleus without collision. 4 lead to the ejection of 4 particles, 42 to the ejection of 3 particles, 82 to the ejection of 2 particles, 43 to the ejection of one particle, and 1 to capture with no ejected particles. In all cases when particles are ejected, the consequent compound nucleus has a mean excitation energy of about 25 to 40 Mev.

Calculations at lower incident energies[21] indicate a transition between 30 Mev and 20 Mev from a situation in which each capture usually leads to a direct ejection to a situation where each capture does not usually lead to a direct ejection. Such a transition is expected from the following considerations. At lower energies, there are three tendencies. One of these is the increase in the nucleon mean free path, which tends to increase the direct ejection process. On the other hand, the number of collisions in which direct ejection is energetically possible falls sharply with decreasing energy. Finally surface reflection effects become important[21,22] so that, even if a particle reaches the surface with enough energy to escape, it is probable that it will be reflected back into the nucleus. The latter two effects easily compensate the first one, so that only a few direct ejection processes occur below 20 Mev or so incident energy on a medium sized nucleus.

3(b) Surface Reactions at Energies $\lesssim$20 Mev

Certain low-energy ($E<25$ Mev) reactions have been given an attention quite out of proportion to their small cross sections. These reactions can be generally called "surface reactions" because they only occur when the bombarding particles pass through the edge of the target nucleus. The usual direct ejection cross sections fall to negligible values at low bombarding energies, largely because of surface reflection. If the incident particle interacts with a target particle while it is actually in the surface, then this reflection effect is not so important. On the other hand, the proba-

[18] R. Serber, Phys. Rev. 72, 1114 (1947); M. L. Goldberger, Phys. Rev. 74, 1269 (1948).

** It may happen that so much energy is left in restricted regions of the compound nucleus that anomalous effects such as "spot boiling" occur. Theoretical work on this phenomenon indicates that it should become important when local nuclear temperatures exceed 7 Mev [see A. Kind and G. Patergnani, Nuovo cimento 11, 106 (1954)].

†† The decay of the compound nucleus is expected to be essentially that pictured by the earlier "strong absorption" theory with particle spectra of Maxwellian type. The fact that the nucleon mean free path is longer than implied by this theory is only expected to affect the spectra to the extent of imposing certain mild modulations on them. These come from the fact that the transmission probability of a nucleon escaping across the nuclear surface is modulated because of its motion in the average nuclear potential.

[19] F. Mandl and T. H. R. Skyrme, Proc. Phys. Soc. (London) A65, 107 (1952).

[20] J. W. Meadows, Phys. Rev. 98, 744 (1955).

[21] Hayakawa, Kawai, and Kikuchi, Progr. Theoret. Phys. 13, 415 (1955).

[22] L. R. B. Elton and L. C. Gomes, Phys. Rev. 105, 1027 (1957); G. Brown and H. Muirhead, Phil. Mag. (to be published).

bility of finding a particle in the surface relative to finding it in the volume is small so the cross sections for these processes are small. The reason why they have excited so much attention is that they can be made to show up very dramatically under suitable experimental conditions. Because of the direct nature of these processes, they tend to produce particles of high energies‡‡ directed in the forward hemisphere. This is in contrast to the compound nucleus process which tends to produce particles of low energy distributed over all angles.

The two main types of surface reaction initiated by incident nucleons are the "surface knock-out"[23] (in which inelastic nucleons appear) and the "surface pickup" (in which deuterons appear). In other words, the interaction of the incident nucleon with a target nucleon may either lead to the escape of one of these two nucleons, or to the escape of both joined together as a deuteron. The so-called "stripping" processes[24] are simply the time inverses of the pickup processes. It is to be expected that the only significant contributions to surface reactions come from the first collisions so that, in calculations, one does not have to follow cascades using the classical statistical method. Instead it is simple to make a quantum mechanical calculation in which the initial and final nuclei are represented as proper quantum states. For a first orientation one can carry out a Born approximation calculation in which the initial and final particle states are represented by plane waves. However this leads to an unwanted and spurious volume contribution[24] unless one simply ignores the part of the Born approximation matrix element from inside the nuclear volume. A better and more consistent treatment consists in representing the initial and final particle states by distorted waves corresponding to some complex potential.[25] §§ In this way, the volume contribution is naturally made small as a result of surface reflection and the absorption inside the potential.

The main qualitative features of surface reactions leading to definite final states are that the details of the angular distribution are sensitive to the amount of angular momentum transferred in the reaction, while the absolute cross section depends on the probability of finding nucleons in the surface of the target nucleus. Consequently, a proper analysis of the data enables one to derive useful information about the spins of states and their reduced widths[26] for break-up into a nucleon and a daughter state.

In general, the cross section for the production of any particular final state at energies $\lesssim 20$ Mev contains two contributions, one from the (delayed) compound nucleus process and one from the (direct) surface reaction. If the incident energy is spread over a range larger than the resonance widths, it can be argued that the two contributions are incoherent. For more sharply defined energies, the two contributions may interfere coherently. As yet, no satisfactory theory that includes both processes in a natural way has been formulated, although some progress has been made towards such a theory.[27]

4. GENERALIZATIONS AND CONCLUDING REMARKS

Hitherto we have consistently retained two restrictions specified in the opening paragraphs, *viz.*, we have taken the bombarding particle to be a nucleon, and assumed the target nucleus to be spherical. We end by removing these restrictions to some extent.

Bombardment with Composite Particles

The scattering and absorption cross sections for bombardments with deuterons, alphas, and heavy ions have been analyzed with a complex potential just like those for nucleon bombardment.[28] One can find complex potentials that fit the data rather well, but the physical interpretation of these potentials is not obvious. One may guess that the absorption coefficient K should be considerably larger than that for nucleon bombardment at corresponding energies since the interaction of any single one of the incoming nucleons with the target will be enough to cause loss of energy. This speculation is confirmed by observation. The experimental values of the mean free paths for alphas and deuterons[28] are much less than nuclear radii and the mean free path for a nitrogen ion is only a few percent of a nuclear radius.[29]

Nonspherical Target Nuclei

It is now known that many target nuclei have a spheroidal deformation in their shape. Although these deformations are not usually large enough to invalidate our discussion of reactions with spherical nuclei, they do give rise to an important new type of process. This is a special and interesting process which cannot be labeled as diffraction or absorption in the senses in

‡‡ The energy spectrum actually depends on a second factor besides the kinematical, momentum conserving, factor which favors high energies. This second factor is the "parentage overlap" between the initial and final nuclear states. Generally speaking, the shell-model predicts that this overlap should decrease with increasing energy separation between the initial and final states. [See A. M. Lane and D. H. Wilkinson, Phys. Rev. **97**, 1199 (1955)]. This factor is somewhat compensated by the fact that the exponential tails of the nucleon wave function are more extended for higher excited states, but, on the whole, it appears that this factor also favors high energies.

23 Austern, Butler, and McManus, Phys. Rev. **92**, 350 (1953).

24 S. T. Butler, Proc. Roy. Soc. (London) **A208**, 559 (1951); P. B. Daitch and J. B. French, Phys. Rev. **87**, 900 (1951); Bhatia, Huang, Huby, and Newns, Phil. Mag. **43**, 485 (1952).

25 J. Horowitz and A. M. L. Messiah, J. phys. radium **14**, 695 (1953); W. Tobocman and M. M. Kalos, Phys. Rev. **97**, 132 (1955).

§§ This procedure is only appropriate if the incident energy covers several states of the compound nucleus. Otherwise the complex potential model loses its meaning [see Sec. 2(d)].

26 J. E. Bowcock, Proc. Phys. Soc. (London) **A68**, 512 (1955).

27 R. G. Thomas, Phys. Rev. **100**, 25 (1955); C. Bloch (to be published).

28 C. E. Porter, Phys. Rev. **99**, 1400 (1955).

29 C. E. Porter, Phys. Rev. **103**, 674 (1956).

which we have used the terms. It consists in the target nucleus being set into rotational motion as a result of the interaction of the incident nucleon with the deformed potential field of the nucleus. This implies the excitation of rotational states and a consequent loss of energy by the nucleon. In principle, the problem can be handled quantitatively by solving the one-body Schrödinger equation for a nucleon interacting with a deformed one-body potential associated with a body possessing a certain moment of inertia. In practice however there are analytical problems in solving this equation, and, in the reported treatments,[30] various approximations have been made in order to estimate cross sections for the processes.

4(c) Concluding Remarks

During the last ten years, several new and important nuclear reaction phenomena have been established experimentally. These include the mountain resonances in total nucleon cross sections, surface reactions, and knock-out processes. In the present review, we have discussed how the earlier models of nuclear reactions have been revised to give some account of these phenomena. The basic essential revision has been to allow the nucleon mean free path to be comparable with nuclear dimensions, instead of assuming it to be very small as in the earlier "strong absorption" model. The result of this revision is a very satisfying synthesis between the ideas of the Bohr model of the compound nucleus and those of the nuclear shell model. The revised model retains many useful features of the "strong absorption" model. For instance, the theory of the compound nucleus that was part of the older picture is still widely applicable within the framework of the revised model.

There is such a wide measure of agreement between the present model and experiment that it seems unlikely that any further substantial revision will be necessary in the future. In such circumstances, it is natural that considerable theoretical attention[16,17] should have been given to the problem of trying to understand the validity of the model from a fundamental point of view. There are many interesting but difficult questions which arise in such attempts. Most of these however belong to the study of nuclear structure rather than of nuclear reactions.

[30] C. F. Wandel, thesis, Copenhagen (1953) (unpublished); D. M. Brink, Proc. Phys. Soc. (London) **A68**, 994 (1955); S. Hayahawa and S. Yoshida, Proc. Phys. Soc. (London) **A68**, 656 (1955); Progr. Theoret. Phys. **14**, 1 (1955).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 2 APRIL, 1957

Statistical Mechanics: A Survey of Recent Lines of Investigation

LÉON VAN HOVE

Instituut voor Theoretische Fysica der Rijksuniversiteit, Utrecht, Nederland

IN this survey I shall try to describe what are, in my view, some of the main lines along which the work in statistical mechanics has developed in recent years, and what are the difficulties encountered. In doing so I shall leave out of consideration more specialized fields of application of statistical mechanics, like solid state, ferromagnetism, superconductivity, and superfluidity. I shall also concentrate on general problems and broad lines of development rather than on any attempt at surveying the recent literature systematically.

As is well known statistical mechanics aims at establishing relations between the observed properties of large systems composed of many identical particles (gases, liquids, solids) and the mechanical properties of these particles and their interaction. The subject naturally subdivides into two parts, one dealing with the properties of many-particle systems which are in equilibrium in the thermodynamical sense, the second treating systems in nonequilibrium situations and dealing with irreversible processes.

I. STATISTICAL MECHANICS OF SYSTEMS IN EQUILIBRIUM

This is perhaps the best instance of a subject where at first sight everything seems to be known, since the basic equations are formally very simple and are familiar to every student of physics. The free energy F of a system in thermal equilibrium at temperature T (canonical ensemble theory) is, classically first and then quantum-mechanically,

$$\exp(-F/kT)=(\hbar^{3N}N!)^{-1}\int \exp(-H/kT)d\mathbf{r}_1\cdots d\mathbf{r}_N \times d\mathbf{p}_1\cdots d\mathbf{p}_N=\sum_\alpha \exp(-E_\alpha/kT) \quad (1)$$

(N number of particles, H Hamiltonian, E_α eigenvalues of the energy). The value at equilibrium of a quantity A is given by

$$\langle A\rangle=(\hbar^{3N}N!)^{-1}\int A \exp[(F-H)/kT]d\mathbf{r}_1\cdots d\mathbf{p}_N =\sum_\alpha\langle \hat{A}\rangle_\alpha \exp[(F-E_\alpha)/kT]. \quad (2)$$

($\langle A\rangle_\alpha$ expectation value of A in the quantum state α).

The simplicity of these equations is only apparent, as is well known. The difficulties find their origin in the fact that the relevant properties of the systems studied manifest themselves only when these sytsems are sufficiently large, so the above equations have to be used in the case of large N (asymptotic limit $N\rightarrow\infty$) where their formal compactness is completely misleading. This explains why a fully satisfactory derivation of physical predictions on the basis of (1) and (2) could, until now, only be achieved in very few cases. The problems met with in this connection can roughly be classified in three groups of increasing difficulty, to be described in succession hereafter.

Imperfect Systems

We shall understand under imperfect systems the many-particle systems with a Hamiltonian $H=H_0+V$ split into a main term H_0 giving rise through separation of variables to a simple evaluation of (1) or (2) and a small perturbation term V mixing the modes of motion which are independent under H_0 alone. This is the simplest type of situation, since all complications connected with the large value of N enter only through the small perturbation term V. As examples we quote imperfect but still very dilute gases (V is here the intermolecular interaction), crystal lattices (the anharmonic forces play the role of perturbation), the Bloch theory of metals (V is the interaction between conduction electrons and lattice vibrations), impurity problems in solid state. Although such problems seem to reduce at first sight to simple applications of perturbation calculus, they actually may present non-negligible difficulties due to the fact that the perturbation, even when small per particle, is in total made large by the presence of many particles. Perturbation theory must, to be adapted to this circumstance, undergo considerable modifications which until now have been carried out in full detail only for classical dilute gases.[1]

How difficult such perturbation problems may be is illustrated by the example of the imperfect Bose-Einstein gas, where the effect of the intermolecular interaction on the Bose-Einstein condensation phenomenon is still unknown, even for the simplest case where the range of the forces is very small compared with the mean de Broglie wavelength. The solution of this problem, in addition to its intrinsic interest, would probably be quite instructive for our understanding of the superfluidity of liquid helium at low temperature. Similarly, a complete understanding of the physical effects resulting from the interaction between Bloch electrons and lattice vibrations in a metal should be very helpful for the eventual explanation of the mecha-

[1] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley and Sons, Inc., New York, 1940), Chap. 13.

nism of superconductivity. Needless to say, the existing treatments of these problems are still very incomplete.

Liquids and Dense Gases

The problems connected with the liquid state and with the gaseous state at high densities are much more difficult than those related to imperfect systems because the complexities resulting from the large number of degrees of freedom, instead of appearing through a small correction term in the Hamiltonian, now play a dominant role. Neither for the calculation of thermodynamic properties, nor for the prediction of the observable local structure properties (e.g., the pair distribution function) does one have theoretical methods deriving the main part of the results from basically simple calculations and providing an approximation scheme of reasonably fast convergence to deal with the remaining corrections.

The most natural approximation method, initiated by Lennard-Jones and Devonshire,[2] pictures each particle of the fluid as moving in the average field of force of the neighboring particles, thus neglecting all correlation effects. Much work has been devoted to the improvement of this so-called cell theory of liquids by including interparticle correlations in various ways. Although, as shown by de Boer,[3] the introduction of correlations involving more and more particles can be constructed as a systematic scheme of successive approximations, the convergence is too poor to make this approach really successful. Another approach, first proposed by Kirkwood and taken over with minor modifications by Born and Green,[4] postulates a simple expression of the three-particle correlation in terms of two-body correlations. Beyond lacking any basic justification for the densities of interest, this approximation turns out to be quantitatively rather poor and it does not seem to lend itself to convenient inclusion of further corrections.[5]

It seems fair to say that no truly satisfactory theory of liquids has been found so far. The reason for this failure can be formulated by saying that liquid structure has not revealed any clear-cut element of simplicity similar to the scarcity of intermolecular collisions in dilute gases, or to the smallness of anharmonic forces in the crystalline state. Still it is not obvious that such elements of simplicity are altogether absent. Thus, despite the complicated n-body correlations present in liquids (with n going perhaps up to $n \gtrsim 10$), there is the striking fact that all correlation effects are of short range (except in the immediate neighborhood of the critical point), a simplifying feature which one might expect to be quite significant. Whether it can form the basis of a more fruitful theoretical approach is not known.

Phase Transitions

While the structure of liquids and dense gases confronts us with the problems just discussed, the occurrence of a discontinuous phase transition between these two states of matter for all temperatures below a certain critical temperature is an even more striking phenomenon to be accounted for. In view of its great generality, this condensation phenomenon must be explainable in terms of the qualitative properties of the intermolecular force (strongly repulsive at short distance, attractive at larger separations beyond which it rapidly converges to zero). Considerable work has been devoted in the last twenty years to the statistical mechanics of the gas-liquid transition and of similar transition phenomena (order-disorder transition in binary alloys, ferromagnetism, etc.), in an attempt to derive the existence and exact nature of the transition from first principles without recourse to *ad hoc* thermodynamical arguments.

Indeed, if arguments of the latter sort are admitted, the theoretical prediction of phase transitions is nowadays a simple matter, and such a half-phenomenological approach has met with repeated success for various kinds of systems, starting with the pioneer work of van der Waals who derived the liquid-gas transition and the existence of a critical point from the simplified equation of state bearing his name. This equation of state gives below the critical temperature an analytical $p-v$ curve (p pressure, v volume per particle) with $\partial p/\partial v>0$ over an interval of v values, from which one then obtains by application of the Maxwell rule the observed non-analytical $p-v$ curve with a horizontal part ($p=$constant) outside of which the inequality $\partial p/\partial v<0$ holds. The Maxwell rule is in this case the *ad hoc* thermodynamical argument. The condensation problem to be discussed here consists in establishing without any such argument the occurrence of the observed phase transition. Relevant to its understanding is the fact that, from the principles of canonical ensemble theory, the inequality $\partial p/\partial v \leqslant 0$ is bound to hold (implying the equivalence of the canonical and grand canonical ensemble theories) as a consequence of the general shape of the intermolecular potential.[6] From a calculation of the sum of states (1) condensation must be found as a $p=$constant part in the $p-v$ curve, i.e., as analytical singularities in the thermodynamical functions. Since for finite number N of particles the expression (1) is a perfectly regular function of the total volume Nv, we must conclude that the abovementioned singularities can only appear when the limiting process $N\rightarrow\infty$ is carried out. It will have to be found from a proper treatment of this difficult limiting case.

Despite many attempts nobody has however suc-

---

[2] J. E. Lennard-Jones and A. F. Devonshire, Proc. Roy. Soc. (London) A163, 63 (1937); *ibid.*, A164, 1 (1938).

[3] J. de Boer, Physica 20, 655 (1954).

[4] J. G. Kirkwood, J. Chem. Phys. 3, 300 (1935); M. Born and H. S. Green, Proc. Roy. Soc. (London) A188, 10 (1946).

[5] For further details on the theories of the liquid state, see J. de Boer, Proc. Roy. Soc. (London) A215, 4 (1952).

[6] L. Van Hove, Physica 15, 951 (1949).

ceeded in giving a mathematical treatment of the problem powerful enough to decide whether the canonical theory is able to account for the condensation phenomenon of gases. Still, very strong evidence that the question has to be answered in the affirmative was obtained from an exact study of a simplified two-dimensional model by Onsager.[7] This study, developed by Onsager for a model of ferromagnetism (the Ising model), has been extended by Yang and Lee[8] to a related model for a gas (the lattice gas). The result is that below a finite critical temperature condensation occurs with a p= constant part in the $p-v$ curve. In the absence of any conclusive treatment of realistic cases, this result is of the utmost importance because none of the over-simplified features of the model seems likely to affect the significance of the conclusion.

Onsager's contribution stands alone, with its well-established positive result, in contrast to a variety of essentially unsuccessful attempts at settling the condensation problem for realistic systems. We shall mention one of these attempts here because much more work has been devoted to it in the last twenty years than to any other. It is the approach initiated by Mayer[9] and designed to derive the condensation point of a gas from the existence of analytic singularities in the functions expressing p and v^{-1} in terms of the activity z. For a dilute gas these functions are obtainable as power expansions in z, with coefficients readily expressed in terms of the intermolecular potential. The hope was that the analytic functions defined by these series, while representing correctly pressure and density for the whole region of existence of the gas, would exhibit a singularity at the condensation point (low density end of the p= constant interval in the $p-v$ curve). Up to now all attempts at establishing these conjectures have however failed: neither could the occurrence of such a singularity be established, nor could one show that such a singularity, if found, would correspond to the condensation point. As to the latter question, a recent result concerning Bose-Einstein condensation shows that a negative answer is perhaps not so completely unlikely as was originally thought. For the ideal Bose gas, while the series expansions of p and v^{-1} in powers of z have long been known to exhibit a singularity at the condensation point, one now knows that the expansion of p in powers of the density v^{-1} (virial expansion) has a radius of convergence that, although finite, is much larger than the value of v^{-1} at condensation.[10] It is of course not excluded that for a classical imperfect gas a similar situation might hold not only for the virial expansion but alse for the expansion in z.

In conclusion, despite extensive investigations by many authors, there is very little progress to report on the statistical mechanics of transition phenomena. This clearly shows how difficult the problem is and underlines the practical importance of the less ambitious half-phenomenological approaches (as the van der Waals theory of condensation and the quasi-chemical approximation for order-disorder phenomena) relying on a thermodynamical argument of the type of Maxwell's rule. It is therefore good to clarify a little more the position of such approaches with respect to the rigorous canonical ensemble theory. They all have in common one aspect which accounts very simply for the partial significance of their results. The approximations which they introduce are always based on the assumption that the system is homogeneous in space; in other words, only homogeneous configurations (i.e., configurations of constant density for the condensation problem, configurations of constant concentration for the order-disorder problem) are taken into account in the evaluation of the sum of states. This approximation, while reasonable outside the transition region, restricts in this region the choice of configurations to unstable ones (possibly metastable near both ends of the transition region) and excludes the infinitely more probable two phase configurations. This explains why under such an approximation the calculated thermodynamic quantities exhibit no singularity and why the observed quantities can be obtained from the calculated ones by thermodynamical criteria of stability. This may however also suggest the possibility of improved approximate discussions of the canonical sum of states which, although less ambitious than the direct attacks on the condensation problem attempted until now, would lead naturally to the observed singularities in the thermodynamic functions and would be such as to insure that the nature of these singularities would not be affected by the inclusion of further corrections.

II. STATISTICAL MECHANICS OF NON-EQUILIBRIUM PROCESSES

In contrast to the case of thermodynamical equilibrium, no general set of equations is known to describe the behavior of many-particle systems whenever their state is different from the equilibrium state and, in view of the unlimited diversity of possible nonequilibrium situations, the existence of such a set of equations seems rather doubtful. Still there are at least two broad questions concerning nonequilibrium states which seem precise enough to be answerable: (i) under which conditions does a many-particle system move toward thermodynamical equilibrium from what may reasonably be called almost every nonequilibrium state, and (ii) by which equations can one describe this tendency toward equilibrium under properly specialized assumptions on the initial state (taking possibly into account the presence of external constraints preventing the system from attaining equilibrium). These are the questions dealt with in the theory of nonequilibrium (or irreversible) processes. Although complete answers

[7] L. Onsager, Phys. Rev. 65, 117 (1944).
[8] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 and 410 (1952).
[9] J. E. Mayer, J. Chem. Phys. 5, 67 (1937). See also B. Kahn, "On the theory of the equation of state," thesis, Utrecht (1938).
[10] W. H. J. Fuchs, J. Rational Mech. Anal. 4, 647 (1955).

still seem very far out of sight, considerable progress has been booked in recent years and the prospect for further development is hopeful.

The recent contributions deal mainly with the second question, but they provide at the same time special answers to problem (i). As is well known, from the desire to solve the latter problem in entirely general terms has grown the whole body of ergodic theory, resulting some twenty-five years ago in the establishment of ergodic theorems (von Neumann, Birkhoff, E. Hopf) of great mathematical scope.[11] Such theorems however suffer from the very generality of the conditions they state for ergodicity, inasmuch as the validity of these conditions for actual physical systems is overwhelmingly difficult to establish and is still undecided. This line of work has therefore advanced little in recent years.[12]

Turning now to less general investigations, we have to mention first the development by Onsager[13] of a phenomenological theory describing in macroscopic terms the irreversible behavior of systems for not too large deviations from equilibrium. This theory plays for irreversible processes a role analogous to that played by ordinary thermodynamics for equilibrium systems and it is ordinarily referred to as the thermodynamics of irreversible processes. Lately it has given rise to a large number of applications and to a critical review of the statistical foundations.[14]

In the last ten years a strong revival of interest is to be noted for the investigation of statistical transport equations, i.e., equations describing the time evolution of some statistical property of the system studied as it moves toward equilibrium. The oldest and best known of these equations is the Boltzmann equation describing how the one-particle distribution function f of a dilute gas varies as a result of interparticle collisions. In familiar notation[15] and for the case of a system homogeneous in space, without external forces, this equation reads

$$\frac{\partial f}{\partial t}+\int (ff_1-f'f_1')ad\mathbf{c}_1d\omega=0, \tag{3}$$

where the last term is the so-called collision term. Equation (3) is valid to lowest order in the density $\rho=v^{-1}$ (v volume per particle) and is accordingly restricted to the effect of binary collisions. Both the derivation of this equation and its extension to higher orders in the density have been the object of various investigations. Much clarification has been achieved on the main difficulty involved in deriving (3) from the reversible Newtonian equations of motion (or from the Liouville equation describing the reversible time evolution of a probability distribution for the total system in its $6N$-dimensional phase space), namely, the need for an exact formulation and proper justification of the assumption being made on the states of the system and accounting for the irreversible nature of the equation to be established (for Boltzmann this assumption was the Stosszahlansatz).[16]

On the other hand a unique procedure has been proposed by Bogoliubov to extend Eq. (3) to general order in the density.[17] It is based on a far-reaching assumption concerning the interparticle correlations in the gas, namely, that they are time-independent functionals of the one-particle distribution function, so that their time variation results entirely from the time variation of the latter. This assumption is supposed to hold after some initial time interval short compared to the time needed to reach equilibrium. With its help and with an additional assumption on the vanishing of correlations in the remote past Bogoliubov obtains a unique prescription to calculate the time evolution of the one-particle distribution function f to all orders in the density. The result is of the form

$$\frac{\partial f(t)}{\partial t}+A[f(t)]=0,$$

where the time-independent functional A is found as a power series in the density. In lowest order Eq. (3) is obtained.

Another approach to the same problem has been adopted by M. S. Green.[18] It is based on a very elegant formal solution (in powers of the density) of the infinite set of equations following from the Liouville equation for the one-particle, two-particle, etc., distribution functions (the so-called hierarchy). An analysis of this formal solution enables Green to investigate in detail the significance and validity of Bogoliubov's assumptions.

Another type of statistical transport equation, also frequently used and especially well adapted to quantum theory, is the equation describing the irreversible time evolution of the probability distribution of the whole system over a complete set of states. Let us consider (compare with the imperfect systems of part I) a quantum system with a Hamiltonian $H=H_0+V$ containing a perturbation term V to be held responsible for the irreversible behavior. The equation in question, now

[11] For a discussion of these theorems and their significance for statistical mechanics, see L. Rosenfeld, Acta Phys. Polon. **14**, 3 (1955).

[12] For a critical study of the quantum-mechanical ergodic theorem see the recent paper of M. Fierz, Helv. Phys. Acta **28**, 705 (1955).

[13] L. Onsager, Phys. Rev. **37**, 405 (1931) and **38**, 2265 (1931).

[14] Two monographs on the subject are I. Prigogine, *Etude Thermodynamique des Processes Irréversibles* (Dunod, Paris, and Desoer, Liège, 1947), and S. R. de Groot, *Thermodynamics of Irreversible Processes* (North-Holland Publishing Company, Amsterdam, 1951).

[15] See, for example, D. ter Haar, *Elements of Statistical Mechanics* (Rinehart Publishing Corporation, New York, 1954), p. 38, Eq. (2.507).

[16] See on this subject the Proceedings of the "Colloque International sur les Phénomènes de Transport en Mécanique Statistique" (Bruxelles, September, 1956).

[17] N. N. Bogoliubov, *Dynamical Problems in Statistical Physics* (in Russian), (Moscow, 1946).

[18] M. S. Green, J. Chem. Phys. **25**, 836 (1956).

often called the master equation, involves the probabilities P_α that the system be in the various stationary states α of the unperturbed part H_0 of the Hamiltonian. It is only valid in lowest order in V and reads

$$\frac{dP_\alpha}{dt} = \sum_\beta (W_{\alpha\beta} P_\beta - W_{\beta\alpha} P_\alpha). \tag{4}$$

$W_{\alpha\beta}$ is a transition probability of second order in V. Equation (4) was first derived by Pauli[19] on the assumption that the phases of the wave function in the unperturbed representation α are randomly distributed at all times. It was recently realized that the perturbations occurring in the many-particle systems of quantum statistics (as examples see the imperfect systems mentioned in part I) possess remarkable analytical properties which essentially result from the systems being very large and the perturbations coupling together their many degrees of freedom. On the basis of this recognition it was possible to give a new derivation of Eq. (4), requiring randomness of phases at the initial time only, and to generalize (4) to arbitrary order in the perturbation. The extended master equation thus obtained is integrodifferential in the time and corresponds to a non-Markoffian stochastic process.[20]

The classical analog of the quantum-mechanical master equation (4) has also been considered. Prigogine and Brout, using action and angle variables, have expressed classically the special properties of the perturbation which had been recognized in the quantum case as being responsible for the irreversible motion. They have thereby achieved in lowest order of the perturbation a satisfactory derivation of a classical master equation.[21] Brout has derived in lowest order in the density a master equation for the irreversible motion of a dilute gas.[22] The relation of such an equation to the Boltzmann equation (3) had been previously clarified by an investigation of Kac using a simplified model.[23]

It is clear from this review of recent work that the statistical theory of irreversible processes is in a period of rapid development. The results attained are undoubtedly fragmentary, but the techniques used to derive them are far from being fully exploited and further progress is to be expected. As a problem which still presents considerable difficulty we might mention the treatment of situations where external constraints (like temperature gradients, concentration gradients, etc.) prevent the system under study from reaching equilibrium. For instance the statistical foundation of the phenomenological equations known to describe nonequilibrium steady state processes has not been sufficiently clarified. Another much more ambitious question is the statistical mechanics of nonequilibrium processes in liquids and dense gases. On this last subject we are greatly hampered by the absence of formal techniques properly adapted to the liquid state (see part I), although there are indications that at least for slow irreversible changes this difficulty might not be so essential.[24]

---

[19] W. Pauli, *Festschrift zum 60 Geburtstage A. Sommerfelds*, (Hirzel, Leipzig, 1928), p. 30.

[20] L. Van Hove, Physica **21**, 517 (1955) reference 16 and Physica (to be published).

[21] R. Brout and I. Prigogine, Physica **22**, 621 (1956).

[22] R. Brout, Physica **22**, 509 (1956).

[23] M. Kac, "Foundations of kinetic theory," Proceedings Berkeley Symposium on Mathematical Statistics (1954).

[24] M. S. Green, J. Chem. Phys. **20**, 1281 (1952) and **22**, 398 (1954); N. G. van Kampen, Physica **20**, 603 (1954).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 2 APRIL, 1957

Superfluidity and Superconductivity

R. P. FEYNMAN

Norman Bridge Laboratory of Physics, California Institute of Technology, Pasadena, California

I AM sorry that Professor Landau was unable to come for two personal reasons. The first is that I have worked on this problem of helium upon which he has also done so much, and I would have liked to have spoken with him about it. The other reason is that I wouldn't have had to give this lecture.

Quantum mechanics was developed in 1926, and in the following decade it was rapidly applied to all kinds of phenomena with an enormous qualitative success. The theories of metals, other solids, liquids, chemistry, etc. came out very well. But as we continued to advance the frontiers of knowledge, we left behind two cities under siege which were completely surrounded by knowledge although they themselves remained isolated and unassailable. The first of the two very similar phenomena for which we still lacked a qualitative explanation is the superfluidity of helium. I may remind you that the discovery that helium flows without resistance through very thin tubes was made as far back as 1911. The second phenomenon is the superconductivity of metals. The fact that electricity flows through some metals without any resistance at low temperatures was discovered, I believe, in 1905. Of these two phenomena, I think we now understand qualitatively the superfluidity of helium, but we do not yet understand qualitatively the superconductivity. I propose to give a brief summary of views on liquid helium insofar as they may give some clue as to the kind of thing that is involved in superconductivity. It is clear, however, that the solution of the superfluidity problem does not give us a very *good* clue because we have solved the first problem but not the second.

First, I make some semiphilosophical remarks about the kind of problem facing us, and the kind of view that I take toward it. I do not want to discuss all of the aspects of superconductors and of superfluids. I want to discuss only the interesting qualitative features, i.e., the curious problem of how does it work more or less. In other words, we would like to make an analogy between the problem of, say, superconductivity and the problem of friction. After all, how is friction explained on the basis of the Schrödinger equation? No one has ever computed the coefficient of friction of two blocks of copper, but qualitatively we feel that somehow when we rub the atoms together we can understand, more or less, how friction arises. But we do not understand, more or less, how superconductivity works and I would like to address my attention to this problem of understanding it *more or less*, not of understanding the details of a lot of special phenomena. In other words, I would like to concentrate here on the problem of interpretation from first principles. We would like to connect the Schrödinger equation directly to some experimental facts. There are many interesting things that happen when one works on a problem in this way. For example, if one makes some approximations, they cannot be justified by comparing them with experiments. It is necessary to justify the approximations directly in terms of arguments from the equations as a mathematical problem.

In the first stage of the development of such explanations there is an enormous amount of phenomenological explanation made which is of great value, and I don't mean to say anything against it. That is just not what I want to speak about. This phenomenological development is vital in combining various experiments together and is helpful in giving us hints as to what to try to explain. If we were any good, however, we wouldn't need the experiments and we wouldn't need the hints. We could simply calculate everything directly from the Schrödinger equation. So I will assume some familiarity with the wide range of experimental properties and will only make a very brief mention of the salient points.

First we will discuss the properties of liquid helium. As a function of temperature, the specific heat appears something like Fig. 1 experimentally. The details near the transition are unknown. There is a certain temperature at which there is a transition, the so-called λ point, and the properties of the liquid below the transition are very peculiar. First, helium remains a liquid all the way down to absolute zero, which we know is due to the zero point motion of the atoms. The reason for the existence of the transition is undoubtedly the same as for the transition in the corresponding Einstein-Bose gas neglecting interaction between the atoms. The Einstein-Bose transition for an ideal gas, however, has quite a different shape. It looks something like Fig. 2, and the problem of why the actual transition looks like

FIG. 1. Specific heat of liquid helium as a function of temperature.

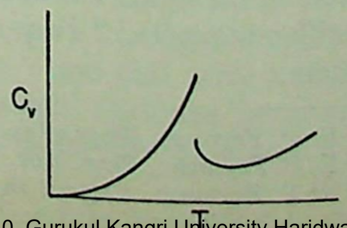
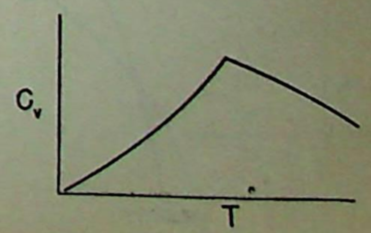


FIG. 2. Einstein-Bose transition for an ideal gas.



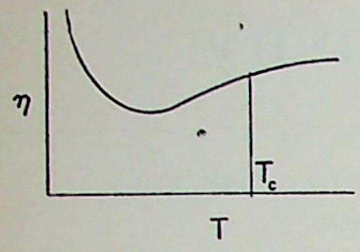


FIG. 3. Coefficient of viscosity of liquid helium *versus* temperature as determined by rotating two concentric cylinders relative to one another.

Fig. 1 when there are forces involved is not yet completely solved. The detailed behavior near the transition, however, is a very delicate and difficult proposition. One is convinced that the transition is due to an Einstein-Bose condensation, but exactly how interatomic forces modify the Einstein-Bose specific heat is not yet thoroughly understood.

To discuss superfluidity, I would like to discuss the properties of helium near 0°K, far from the transition region. If one measures the viscosity of the liquid as a function of temperature by rotating two concentric cylinders relative to one another, the coefficient of viscosity behaves as in Fig. 3. Near the transition temperature nothing in particular happens.

On the other hand, if one measures the viscosity by making the liquid flow through a very narrow tube, the liquid flows through such a tube with an apparent viscosity above the transition temperature, but below the critical temperature it flows with apparently no viscosity. Incidentally, if one measures the resistant force to push the liquid through a tube at different velocities then above a certain critical velocity a resistance does arise but below the critical velocity the resistance is zero. A relatively thick tube has a lower critical velocity than a thinner tube as is shown approximately in Fig. 4. The explanation of these things at least in a qualitative way, has come essentially from Landau. He pointed out that we should imagine that the liquid at, say, absolute zero is a perfect fluid for some reason (or rather it is a perfect liquid when the flow is potential flow and has zero curl). As we heat the sample up, the energy goes into local excitations of some kind, for example, quantized sound waves called phonons and excitations of higher momentum and energy called rotons. If, as the temperature rises, the excitations are few and far between (a fact directly indicated by the very small specific heat), then they will be localized inside the liquid and bounce around within it. The local excitations behave very much like a gas, and the system behaves in many respects like a mixture of two fluids. Thus we have a basis for the two-fluid model, which was a phenomenological model.

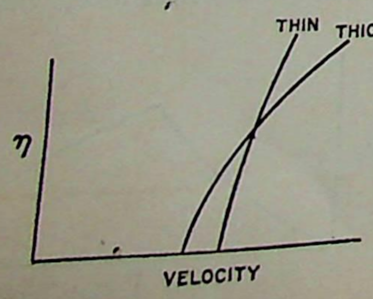


FIG. 4. Coefficient of viscosity *versus* velocity in thin and thick tubes.

The energy of the excitations required as a function of their momenta to produce the specific heat and general properties seems to be a curve of the general nature of Fig. 5. The curve is linear for smaller momentum, meaning that at long wavelengths sound waves are excited. But the states of higher momentum excitations are of some other kind. At any finite temperature one excites states mainly in two regions; small p near where the curve is nearly straight, and near the bottom of the concave part of the curve. The lower states are called phonons and the higher states are called rotons.

The reason why the flow is perfect, from the point of view of Landau, can be seen in the following way. Put a little ball inside the liquid at absolute zero and try to make it move through the material, allowing the liquid to flow around it. How can the ball lose energy to the liquid? Only by exciting states inside the liquid, and if the aforementioned states are the only ones available, it is easy to show by the conservation of energy and momentum that a very high velocity is needed before we have the necessary energy at the given momentum transfer to produce such a state. The ball must have a velocity v high enough that the straight line $E=pv$ intersects the curve in Fig. 4. Actually, the

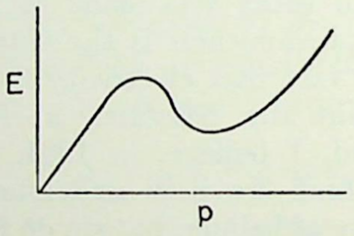


FIG. 5. Energy of excitations required to produce specific heat and general properties of liquid helium in the two fluid model as a function of their momentum.

resistance sets in at a very much lower velocity by a factor of about 100, and this fact has to be explained. We have to explain *not* the superfluidity, but why it is not quite so perfect a superfluid as it should be from this point of view.

The essential feature of the liquid is that there are very few states of low energy, a fact directly shown in the specific heat and one which also explains the superfluidity ultimately. So the question is, why are there so few states of very low energy? Why in this liquid, for example, can't we get a very low energy quantum state, say, by letting some portion of liquid slowly rotate? The portion of liquid could have a big moment of inertia and one unit of angular momentum would have a very low energy. Or alternatively, say, one of the particles inside the liquid decides to drift slowly through the liquid while the others remain stationary. The explanation of why there are no such low energy states in the liquid has been worked out and published in different places.[1–3] Time does not allow me to describe it here.

[1] R. P. Feynman, Phys. Rev. **91**, 1301 (1953).
[2] R. P. Feynman, Phys. Rev. **94**, 262 (1954).
[3] R. P. Feynman, *Progress in Low temperature Physics* (North Holland Publishing Company, Amsterdam), Vol. I, p. 17.

But there is a point that isn't discussed very often: the question of how is it that we do get some resistance even at a velocity lower than the velocity required to create an excitation directly. We must discuss the character of the wave functions representing the liquid in various states of flow. If the wave function for the ground state of the liquid is represented simply by Φ, then it is very easy to represent the same system in motion at a uniform velocity in a straight line. The wave function for the moving system is simply

$$\Phi \exp(i\mathbf{P}\cdot\mathbf{R}_0/\hbar)=\Phi \exp(im\textstyle\sum_i \mathbf{V}\cdot\mathbf{R}_i/\hbar), \qquad (1)$$

where $\mathbf{P}=Nm\mathbf{V}$ is the momentum of the whole liquid of N atoms moving with velocity $\mathbf{V}$, and $\mathbf{R}_0=\sum_i\mathbf{R}_i/N$ is the position of the center of gravity of the atoms whose individual positions are $\mathbf{R}_i$. This wave function represents a whole system moving at a uniform velocity, but we would like to represent something more complicated. Suppose the liquid flows in some more elaborate fashion in which the velocity is not the same everywhere. If the velocity is a very slowly varying function of position $\mathbf{V}(\mathbf{R})$, and does not change appreciably in an atomic distance, we can try to represent the wave function by simply making the velocity of the previous formula dependent on position. The wave function now becomes

$$\Phi \exp(im\textstyle\sum_i\mathbf{V}(\mathbf{R}_i)\cdot\mathbf{R}_i/\hbar)$$

which has the mathematical form

$$\Psi=\Phi \exp i\textstyle\sum_i S(\mathbf{R}_i)/\hbar$$

where $S(\mathbf{R})$ is some function of position. We have said it is $m\mathbf{V}(\mathbf{R})\cdot\mathbf{R}$ but this is not quite right. The momentum for a wave of slowly varying wavelength is the gradient of the phase, not the phase divided by the coordinate, so the correct relation is

$$\mathbf{V}=\nabla S/m.$$

This implies that the curl of the velocity is zero. We have thus only accomplished the description of flow which has zero vorticity by such a method, and the problem of how to describe the flow of the liquid with vorticity becomes very serious. There is no structure in the liquid with which we can tie down the atoms in any given region; they are all the same and identical.

I have already published some work on this problem and Onsager eight years previously[4] also published the same type of discussion. So for variety I'll give a slightly different argument, the first one that I thought of. It is not a good one but it gives us a clue as to what's going to happen. I tried to *force* the liquid to have vorticity, to have some kind of nonpotential flow. I imagined that the liquid in region I of Fig. 6 is flowing to the right with the velocity v and that some kind of barrier or sheet of infinitesimal thickness separates region I from region II. In region II the velocity is zero. In region I the wave function should look something like Eq. (1), and the wave function in II is just Φ. If such a separated flow has been progressing for some time, what happens if all of a sudden I pull out the sheet? It's got to do something. In spite of the fact that the velocity according to formula (2) always has to have the form of a gradient of S, it cannot be a gradient in this case. I have to make a rule that the atoms in the upper section are in a different state than the atoms in the lower section. If I simply restrict the motion so there is no flow across the boundary I will get a possible kind of motion. There must be a layer here with an associated surface tension where the atoms in I can't wander into II or vice versa. Hence, there is a certain energy associated with relative flow with a sheet of surface tension in between.

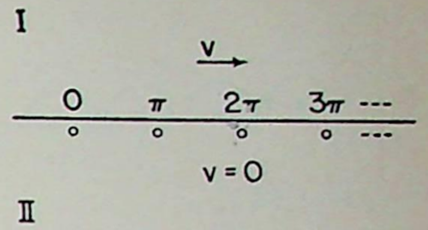


FIG. 6. Sketch showing different fluid motions separated by an infinitesimal sheet.

Suppose that the velocity in I becomes smaller and smaller. Is it reasonable that, even with a very small drift in I, the surface energy per unit area is still the same no matter how small this velocity is? Why can't the atoms from I go into II? If they can't, then I need the surface tension, but if I can allow them to mix, I don't need that surface tension. They can't mix because the wave function must vary from one position to another. Suppose an atom moves along just above the interface of Fig. 6. The phase must change according to formula (1), from 0 to π and so forth as sketched in Fig. 7. At points where the phase is π, the wave function cannot be smoothly connected with the constant wave function in I, while at those points where the phase is zero, or 2π, etc., the wave function can be connected smoothly. The cosine wave and the constant 1 have places where they are equal. In those places I can erase the surface locally without any effect. What I imagine in the next approximation is that there are certain slots of surface tension with flow around the slots. What is the best length of slot? There is a certain energy of motion around the slot, and so on. If I continue the distortion further I probably lower the energy in every case until I come down to something like a string of holes instead of slots. The velocity distribution then is sketched in Fig. 8 and the wave function changes by 2π when it goes around one of the holes, as it went around the slot. But it is now con-

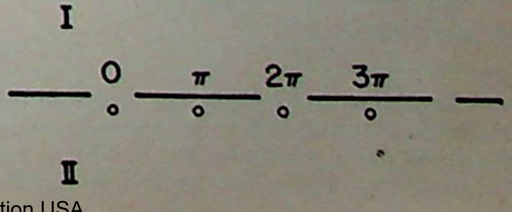


FIG. 7. The sheet of Fig. 6 is erased at those points where the wave functions can be joined smoothly.

[4] L. Onsager, Nuovo cimento Suppl. 2, 6, 249 (1949).

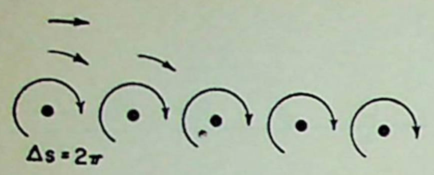


FIG. 8. Velocity distribution when the slots of Fig. 7 distorted into a string of holes.

tinuous everywhere except at the holes. The curl of the velocity is still zero, but S is a multiple-valued function of position (and is not defined in the holes). Thus the integral of the velocity around a circuit is $2\pi n\hbar$. In this way I was led to the idea that when liquid He flows, the vorticity is quantized and all the vorticity is along certain lines of one unit each. And when I was all finished I found a reference to a remark made by Onsager in a meeting in Italy, to a lecture on a different subject, in which he mentions the same solution!

What happens is this. If the object moves too quickly through the liquid vortex lines are formed. If the velocity is too slow there is not enough kinetic energy in the motion to create even one vortex line. For a very thin tube, for example, there is not enough kinetic energy to construct a vortex line. Before a vortex is made, there is no resistance and the whole system coasts together. With thin tubes and velocities that are not too high, we have not enough energy to create even the first line so the material maintains its superfluidity. There remains a theoretical problem to establish these lines firmly as they have not been established very well theoretically so far. The argument used is not very good so it is desirable to make a better theoretical argument about the quantized lines. In the meantime I would like to see an experiment to show that the lines of vortices really exist directly.[5] The resistance, for example, would be irregular when it first starts if it comes in quantized units. How the vortex lines actually form from a surface has defied my analysis. This, with the omission of the part about why there are no excited states, is all that I have to say about helium.

Problems associated with this field still remain. There are no problems of understanding the phenomena qualitatively any more, but there are problems of quantitative understanding. For example, we need more details about the exact wave functions for the rotons and phonons and their properties. We would like to calculate the cross section for the collision of two rotons. The collisions determine, for example, the viscosity and absorption of second sound, as was pointed out by Khalatnikov and others. They use an arbitrary cross section to fit the experiment, but the question is how to calculate the cross section from first principles. A much more interesting problem probably is the behavior —statistical mechanically, or otherwise—very near the transition.

[5] Experiments by Hall and Vinen on the attenuation of second sound in rotating helium give strong evidence for the existence of the lines. H. E. Hall and W. F. Vinen, Conf. de Phys. des basses temperatures, Paris (September, 1955), Paper 22.

Today we do not have a good description of what happens within a few tenths of a degree on either side of the transition. We have approximate views about the shape of the curve, but whether it has a vertical tangent or contains a jump in specific heat is not known.

Another problem is the problem of what is the ground state energy. Here we have a strongly interacting system of particles with known interaction and the problem is to find the energy, or other properties of the system in the ground state. It is a quantum mechanical problem and has never been solved, but it would be interesting to work that out from first principles. The energy of a roton has been worked out now. The curve of energy against momentum has been calculated experimentally with the data and also been computed theoretically. The energy of the roton theoretically is 11.5 deg and experimentally is 9.6 deg. To make the theoretical calculation we had to use some properties of the ground state, namely the correlation function of the ground state. To put everything on purely first principles, it would be necessary to calculate the correlation function of the ground state by solving the Schrödinger equation, and I don't know how to do that.

Another liquid of great interest is liquid He^3. I do not believe that liquid He^3 is a superfluid. I think that there are an enormous number of states reaching to absolute zero. I believe that the specific heat would be proportional to the temperature as it is for an ideal gas and that the forces would not modify it. I have no *reason* to believe all these things; I have thought a lot about He^3, but I find it too difficult to analyze accurately from first principles. If there is a high density of states then He^3 will not be a superfluid in the same sense as He^4. I think that the same thing will happen in nuclear matter, so that superfluid is not a very good term to apply to the nucleus. It is a quantum fluid but I don't think it is a superfluid.

Another interesting problem is the following example from He^3. The particles interact with strong forces in a Fermi system. The problem is to determine the temperature dependence of the viscosity of the system as T goes to zero. There are a large number of similar interesting problems in He^3, and it would be fun to do them before the experiments, for the first time. I don't think anybody has ever computed anything in solid-state physics before the experimental result was out, so we have consistently predicted only what we have observed!

Now we turn to the unsolved problem of superconductivity and again talk from the point of view of what we will have to do from first principles to solve this problem. The first thing that is troublesome is that the metals that become superconducting are always the complicated ones near the middle of the periodic table. Superconductivity depends on the crystal structure, some metals undergo the transition in one crystal form

and not in another. Also, some alloys are superconductive even though neither of the constituents is by itself superconductive. The simplest metals, the alkalis and alkaline earth metals, are not superconductive. The temperature of the transition corresponds to an energy $kT=10^{-4}$ ev, while the electron energies are of the order 10 v or so. Hence, we have a very delicate problem here. It is not like the case of helium, where the energy of the transition is comparable to the thermal energies involved. In the helium case, when the thermal wavelengths, or the deBroglie wavelength corresponding to a temperature of the transition, becomes equal to the spacing of the atoms then all the excitement occurs. On the other hand, in superconductivity the effect is very small and therefore quite subtle. I have to make some excuse, you see, for not having immediately solved the problem.

The specific heat of a superconductor looks something like Fig. 9. According to the theory of an ideal electron gas, the specific heat should be proportional to T. Experimentally, there appears to be a discontinuity in the specific heat at the transition, and the electronic specific heat at very low temperatures may be exponential. The curve is certainly not linear with T; it is much, much less.

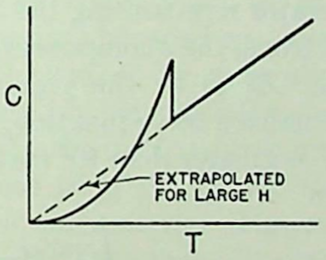


FIG. 9. Sketch of specific heat of a superconductor *versus* temperature.

It is possible to maintain the material in the normal state with a sufficiently large magnetic field. If one applies a magnetic field for different temperatures the stronger magnetic field for lower temperatures, the curve of the transition temperature against magnetic field looks like Fig. 10. Even at absolute zero, the superconductive state can be destroyed by an application of a sufficiently strong magnetic field, and the specific heat of the normal state behaves linearly right down to zero temperature.

The electromagnetic properties of the superconductor have been studied in great detail and summarized in a very simple way by London, at least for low frequencies. London deduced that the law stating that the current density is proportional to the vector potential correctly describes all of the properties. I have to omit all the details of the experiments and the arguments which led London to this law. It is necessary that the vector potential $\mathbf{A}$ be in such a gauge that the normal component of the current density at the surface be zero everywhere. London also gave a kind of explanation of how this may have come about. The formula for the current density of the electrons in the quantum mechanics is given by

$$\mathbf{J}(x)=\frac{e}{m}\sum_{k=1}^{N}\int\Psi^*\left(\frac{\hbar}{i}\nabla_k-\frac{e}{c}\mathbf{A}(x_k)\right)\times\Psi\delta(x-x_k)d^Nx,$$

where Ψ is the wave function of the electrons and the integration is carried out over all electron coordinates. Now suppose that for some reason a small electromagnetic field $\mathbf{A}$ changes the wave function Ψ very little. For the wave function without the potential of course, the mean value of the current would be the expectation value of $\sum_k\delta(x-x_k)(e\hbar/mi)\nabla_k$, alone and would be zero. If the wave function does not change much, this term is still essentially zero. Then all one has is the vector potential at the point x averaged over all the electrons. In other words, one has $\mathbf{J}(x)=-ne^2/mc\mathbf{A}(x)$. London would then explain his law on the assumption that the wave function does not change much when a small magnetic field is applied.

Now what is it that we would like to explain about superconductivity and with what are we going to explain it? There are a whole set of properties such as the specific heat, the London equation, etc. to be explained in terms of the Schrödinger equation. The first question is, which property shall we choose first to try to explain? The next question is whether we need the full Schrödinger equation, or can we take an approximate Hamiltonian that is simpler to work with as a reasonable model of the actual situation.

I would like to maintain a philosophy about this problem which is a little different from usual: It does not make any difference what we explain, as long as we explain some property correctly from first principles. If we start honestly from first principles and make a deduction that such and such a property exists—some property that is different for superconductors than for normal conductors, of course,—then undoubtedly we have our hand on the tail of the tiger because we have got the mechanism of at least one of the properties. If we have it correct we have the clue to the other properties, so it isn't very important which property we explain. Therefore, in making this attempt, the first thing to do is to choose the easiest property to handle with the kind of mathematics that is involved in the Schrödinger equation. I want to summarize some thoughts on this question, although they do not represent a solution. They represent a statement of the problem and a little bit of a personal view.

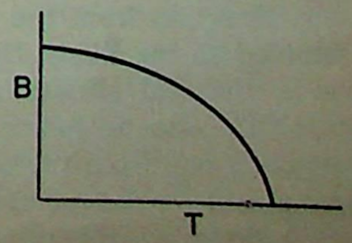


FIG. 10. Transition temperature *versus* magnetic field.

I decided it would be easiest to explain the specific heat rather than the electrical properties. In the case of the electrical properties we have to consider both the temperature and the magnetic field, which is more complicated than the temperature alone. But we do not have to explain the entire specific heat curve; we only have to explain any feature of it, like the existence of a transition, or that the specific heat near absolute zero is less than proportional to T. I chose the latter because being near absolute zero is a much simpler situation than being at any finite temperature. Thus the property we should study is this: why does a superconductor have a specific heat less than T? Let us say that the specific heat is essentially zero. Thus the density of states above the ground state is very much less than that of the ideal gas. It is orders of magnitude of infinity less, to get a different power of T. The property to be explained for superconductors is that the ground state is separated from the higher excited states by a region where the density of states is low. The simplest thing would be that there are no states in between and in fact it may be that that is the case. The specific heat curve looks like it. Let us suppose that the density of states is very low just above the ground state. All we have to do is to explain why the excited states have a definite separation energy from the ground state, or that there are only a very few states near the ground state.

I have not got much further and this conclusion is well known to many others, but I would like to make some remarks on some work in that direction. I tried to state this property directly in terms of some property of the Hamiltonian. First why is the density of the excited states so high in the ideal gas, and how could it fail to be so high in a superconductor?

The ideal gas is easily represented in momentum space by drawing a Fermi sphere with a lattice of momentum that is very fine. The lattice spacing depends on the size of the container and the spacing, ϵ, is inversely proportional to the cube root of the volume of the container. If one calculates the excitation energy of the system at temperature T, one realizes that all the states within a range of kT of the Fermi surface are excited. Thus the number of states excited is kT/ϵ times the number of states around the sphere, which is the square of the radius times $4\pi/\epsilon^2$. All of these states are excited on the average by the energy kT, so this total excitation energy is proportional to the volume and to T^2 so the specific heat is proportional to T. Now where did this enormous specific heat come from? Essentially, it came from the fact that there is a state in any direction from the center of the Fermi sphere. It seemed to me likely that here is the feature that fails in the real situation. How can we define this failure? In order to define one of these states, and to distinguish one from the other, we have to be able to locate points on the sphere. That is to say if we expand our functions in terms of angular variables we must be representing states of enormous angular momentum of very low energy.

Consider a big box, the metal. If one excites a particle, that is, takes it out of one state and puts it somewhere else, then this excitation has a tremendous angular momentum around the center of the box, because it has finite momentum (the Fermi momentum) and is a long distance from the center. It takes practically no energy to create this large angular momentum, according to the ideal gas picture. One hardly has to lift it out of the shell. Thus just above absolute zero, with an infinitesimal energy, one can create very high angular momentum. If the distance required to get superconductivity in small grains corresponds to 10^4 atoms or so, then this means that in a normal conductor in order to create $10^5\hbar$ units of angular momenta requires infinitesimal energy.

Therefore, in the superconductor we probably have the following property. The difference between the energy of the lowest state with large angular momentum and the ground state, must approach a finite number as the volume of the box goes to infinity. This is the property that the superconductor has that I now look for. But the angular momentum of the electrons does not commute with the Hamiltonian. Let L be the operator representing the angular momentum of all the electrons, the z component of which is $\sum_i(x_ip_{iy}-y_ip_{iz})$. Take E_0 to be the ground-state energy. If f is any normalized wave function, for which the expected value of L is greater than 10^5 (the exact value is unimportant), then

$$\int f^*(H-E_0)fdv>\eta,$$

for

$$\int f^*Lfdv>10^5,$$

where η is finite as the volume goes to infinity. This is not true of an ideal gas, but I believe it is true of the Hamiltonian of the superconductor.

I don't know if I have made any progress over the simple remark that the density of states above the ground state is very low. That is as far as I was able to get with the specific heat problem. If one assumes the truth of the previous equation one cannot necessarily explain all the properties of superconductors. But if one can find the *correct* explanation of one of the unique properties of a superconductor, one can then see the mechanism for all the other properties. Actually, we can almost see how the electrical properties will come out of such a procedure. If one considers a cylindrical piece of metal with a field along the axis of the cylinder, then the Hamiltonian of the system with the field is related to the Hamiltonian without the field by $H'=He/c-B\cdot L$. That is, the $A\cdot p$ term is equal to $B\cdot L$ for such a cylinder. Let ϕ be the ground-state wave function. In the ground state the expected value of H is

the lowest energy, E_0. $(\phi,H\phi)=E_0$. The expected value of the angular momentum is zero. For any other state, which has a reasonable amount of L, we get

$$(f,Hf)=E_0+\eta-10^5(\hbar e/c)B.$$

If B is not too big, this energy is larger than E_0, so with a small enough field, we would get the lowest state, which would be the ground state not the new state, and therefore the wave function would be the same as it was before. States with smaller angular momenta do not affect the argument since B has little effect on these. Thus, London's explanation of superconductivity would result. This argument is not very good. It must be more complete, since one has to show that if we took other states where L is larger than 10^5, the energy rises fast enough for a sufficiently large L. When we do finally understand the reason for this gap, we will understand how the energies will vary with the angular momentum of the states that are created.

Now the next part of the problem is with what do we make this explanation? That is, what approximate Hamiltonian should we use? The question is, have we included enough junk in the model to include the effect of superconductivity? The first and simplest model is the model of the ideal gas of electrons. There we know we do not get superconductivity, because we do not get the right specific heat. The next thing to be added that seems to be very important is Coulomb interaction. Sometimes it has been thought that a Coulomb interaction would make some kind of lattice of electrons which would move through the material. However, everyone argues, and I think correctly, that this cannot produce the small effect, because it is too large. That is, the Coulomb interaction energies in a material are of the order of a volt or so per electron, while the energies we are looking for amount to 10^{-8} v per electron. We have all kinds of other effects, such as spin-orbit couplings, spin-spin couplings, magnetic interactions, and so on. Because we have factors v^2/c^2 in these cases, we get the right order of magnitude.

But there is one little piece of experimental information of vital importance, which shows that we will not get superconductivity even if we include all these effects. That fact is that the transition temperature depends upon the ionic mass. Or better, the energy difference between the superconductive state and the normal state, depends upon the mass of the ions. If these v^2/c^2 effects were the cause of superconductivity, it would not depend upon the mass of the ions.

Could it be that the superconducting state depends upon the mass, not because the phenomenon which produces the energy depends upon the mass, but because when you heat the material up, the shaking of the lattice breaks the state down at a temperature that depends upon the mass? I think this approach is wrong. Since the difference in the energy even at absolute zero depends on the mass, I think that the energy must involve dynamically the motion of the lattice. So we have to include lattice interactions. Now the hope is that if we include lattice interactions, we will not have to include these other interactions.

One of the most popular models of a Hamiltonian to explain superconductivity is the one used so much by Frohlich and Bardeen. Their model consists of a jelly of positive charge, an elastic continuum which can vibrate, in which the band structure of the electrons is neglected. In the jelly of positive charge, one considers a gas of electrons of the same density as the charge density of the positive charge. The electrons do not interact with each other, but they all interact with the vibrating jelly. It is the simplest Hamiltonian which has been proposed that would seem to be able to represent superconductivity.

A lot of people have tried this model. However, they are not absolutely satisfied. Some still worry whether the band structure has something vital to do with the phenomenon. Perhaps superconductivity does not occur in the case of sodium etc., because the band is so very simple and it occurs only for the complicated metals, because their band system is complicated. There may be small regions in momentum space, for instance, where the electrons behave as positively charged particles, that is, places where the conductivity is by holes and other regions where they behave normally. There is some indication that this is the case because it has been noticed that the Hall effect is very small when the material has a tendency to be superconductive. The Hall effect is very small when the positive and negative carriers cancel. Thus some people think that this, in conjunction with the lattice vibrations, may have something to do with superconductivity. Of course, that makes the problem more complicated, because it would mean that if Frohlich and Bardeen could solve their model exactly, they still would not find superconductivity, since it would still involve only negative carriers.

As long as one makes a list of such things, one can always add a few things which might have something to do with it. In case they do, at the end then people can say, "Oh, he thought of it," but I don't believe the one I will mention has anything to do with it. The closed shells have been treated adiabatically and forgotten about and they might have something to do with superconductivity. I just do not think that the closed shells have much to do with it.

I also tried the same thing as Frohlich and Bardeen, which is to solve the Frohlich and Bardeen model with such precision that the estimated errors in the calculation are less than the difference in energy one is looking for. I made diagrams and loops. When one is working with partition functions, it turns out there are all kinds of wonderful tricks that one can do with the diagrams. I calculated the specific heat to enormous precision, and the specific heat was still proportional to T. However, a student of mine, Michael Cohen,

pointed out that the series expansion which I was carrying out may not lead to the lowest state at all. Suppose, for example, that one has a harmonic oscillator and considers the effect of a small perturbation, ϵx^3, on the ground-state energy. Regardless of how small ϵ is, one can never derive the answer from perturbation theory even though the series appears to be nicely convergent. The reason, of course, is that if one goes far enough from the origin the potential becomes arbitrarily large and negative, and hence no lowest state exists. In other words, what happens is that one starts with a function which is qualitatively different from the correct function and the right answer lies in a qualitatively different place. Thus when one starts with a simple kind of wave function to make a succession of approximations, one cannot produce the qualitative features of the isolated lowest state of the real system with the original wave functions. Nor can one be sure that the qualitative difference does not exist just because one computes so accurately. Thus, even if one treats this model very carefully and accurately, as long as it is done by a series method one can never be certain that the model does not lead to superconductivity. That is, we may simply be calculating the normal state specific heat very carefully. It is going to be very hard to show that the phenomenon does not exist in this model. It is also going to be hard to show that it *does* exist.

I am now brought to the same position as Casimir, who first told me about this problem. He said, "There is only one way to go about working this out. It is simply to *guess* the quality of the answer." I think I have come unfortunately to the same position, what we must do is not compute anything, but simply *guess* what makes the ground state isolated at a lower energy. That is, guess what kind of correlation exists at long distances. Why haven't we theoretical physicists solved this problem yet? We have no excuse that there are not enough experiments, it has nothing to do with experiments. Our situation is unlike the field, say, of mesons, where we say, perhaps there aren't yet enough clues for even a human mind to figure out what is the pattern. We should not even have to look at the experiments. Every time we look at another experiment, we make the problem easier. It is like looking in the back of the book for the answer, which is slowly being unveiled by the details of the various experiments. There is no reason to require the experiments. The only reason that we cannot do this problem of superconductivity is that we haven't got enough imagination.

Introductory Remarks on Meson Theory

HIDEKI YUKAWA

Research Institute for Fundamental Physics, Kyoto University, Kyoto, Japan

It is really very hard to say something definite and precise about meson theories. This is partly because of the strong interaction between mesons and nucleons, as well as between two nucleons. But more than that, it is so because any simple-minded meson theory cannot be a complete and closed theory. Obviously, it is much more an open theory than quantum electrodynamics. There are two reasons for this: First, the existence of new strange particles compels us to enlarge the scope of the theory so as to include these particles; second, we believe some real fundamental change is necessary in quantum field theory. We do not know when and how it will come, but the future modification in field theory will affect meson theory substantially whereas, in the case of quantum electrodynamics, the change will be much more academic. For these reasons, it is clear that the validity of any simple meson theory is considerably limited.

Despite the fact that the present situation is difficult, definite progress certainly has been made in meson theory in recent years due to Taketani, Levy, Chew, and many others. This has been partly due to the increase in our knowledge of the pion-nucleon interaction. In the early days we had only limited knowledge about the nucleon-nucleon interaction. We actually knew nothing about mesons themselves. They were just introduced in order to account for nuclear forces. At that time I was a little more ambitious; I wanted to incorporate the phenomenon of beta decay into the same scheme, but this was also rather speculative. Though important progress has been made in recent years, there are still many uncertain points in connection with the pion-nucleon and nucleon-nucleon interactions. All we can say now is that we can correlate pion-nucleon scattering and a part of the nucleon-nucleon interaction in a consistent way. Of course, you may still argue whether this could be done really consistently or not, but it is significant that, if we adopt a rather naive approach to meson theory and adopt the pseudovector coupling with cutoff, then we reach the conclusion that the coupling constant and the cut-off momentum necessary to account, semiquantitatively, for pion-nucleon scattering at energies up to about 200 Mev, agree quite well with the corresponding parameters which are necessary in order to explain the nuclear potential between two nucleons at distances larger than 0.6 or 0.7 times the meson Compton wavelength.

There are still many questions unanswered; we cannot yet decide which of the two interactions linear in the meson field, i.e., the pseudovector or the pseudoscalar interaction, is correct, and further, we do not know exactly what is meant by the cutoff in the theory. A safe way of explaining the cutoff is to accept the view that it is a mere expression of our ignorance. Perhaps this is an extreme viewpoint, but it is not easy to express more moderate views clearly. In the case of the pseudoscalar coupling, it is well known that the effect of nucleon pair creation is so enormous that so far there is no satisfactory way of dealing with it. This question is certainly very intimately connected with the question of the nucleon-antinucleon interaction. We do not yet know exactly what the interaction is; perhaps the complete understanding of the antinucleon and nucleon system will come in a later stage of development in what we may call a comprehensive theory.

There was another feature in the simple-minded meson theory from which we started. Twenty years ago, beta decay was a very important and outstanding problem. Although Fermi's theory worked very nicely, we could think at that time that we didn't need a direct Fermi interaction, but could account for beta decay equally well through the meson field, by introducing some weak interaction between the meson field and electron-neutrino system. Later such attempts became more and more difficult for two main reasons; one was that we came to know that the interaction responsible for beta decay is mainly scalar and tensor. This is contradictory to the pseudoscalar nature of the pion. Furthermore, the intervention of the mu-meson spoiled the whole picture. Thus, as for the weak interactions, we could not stay in the meson theory in the narrow sense, but we had to go over immediately to more comprehensive theories including other types of particles.

Now let us go back to the case of the strong meson-nucleon interaction. The assumption of a strong meson-nucleon interaction was generalized in recent years, so as to include strong interactions between other types of particles. A compact scheme which was originated by Pais and later refined by Gell-Mann and independently by Nishijima for strongly interacting particles was reached on these lines. Here one of the important concepts in meson theory is taken over. That is the concept of charge independence, which is reflected in meson theory as the so-called symmetrical coupling. If you expand this concept in such a way that you can include strange new particles, then you arrive at the concept of isospin space which was introduced by Pais, and this concept of isospin

space works quite well, if you consider only strongly interacting particles.

Very recently, Schwinger has suggested that the analogy between the electromagnetic interaction and meson-nucleon interaction could be carried much farther. In that case, the interaction between pi mesons and K mesons may play a very important role. I do not know yet what will be the correct way of interpretation of strongly interacting particles, but it is clear that we need, at least, one more degree of freedom, which is usually called the strangeness. The notion of strangeness is very important in that we cannot hope to construct from any simple meson theory a comprehensive theory including strange particles, just because pions and nucleons are not strange. You need, at least, one type of strange particle in order to have the whole system of particles. This situation can be compared with the old problem of constructing composite particles from more elementary particles; you can hope to construct particles with integer spin from half-integer spin particles, but the opposite is not true. From the standpoint of a composite particle theory in a broad sense, the half-integer spin particles are necessary from the beginning; for the same reason, the strange particles are also necessary. But, of course, there are other standpoints and whatever be our standpoint, we cannot deny that the Gell-Mann-Nishijima scheme is so nicely mapped out for strongly interacting particles, that one of the main questions for future study is, perhaps, "What would be the real significance of these new degrees of freedom?" I shall come back to this subject later on.

The next question to be considered is that of weak interactions. The problem of weak interactions is not yet answered, or, at least, I cannot easily accept the current interpretations, because some of the conservation laws which hold in the case of strong interactions are violated in the case of weak interactions. If we know merely that some conservation laws are just violated, I don't think you can hope to construct a consistent theory, which implies certain universal symmetry properties. If you have started from a theory with perfect symmetry in isospin space, it should stay on until the end. Of course, the violation of selection rules which is connected with isospin and strangeness may not be very serious, because, after all, we do not know much about the significance of these new degrees of freedom; but if the conservation laws which are connected more directly with our ordinary space-time world begin to be in question, the situation is really serious. All of you know that they really have been the subject of vigorous discussions in connection with the tau-theta dilemma. I do not know whether or not parity should be the only such space-time symmetry property in question; because we cannot wholly reject the possibility that angular momentum conservation is also violated in weak interactions.

The intervention of the mu meson completely destroyed the picture which I had had in mind. Today, I still do not understand mu mesons at all. Where does the big difference between the masses of the electron and the mu meson come from, if we suppose that the mu mesons interact with other particles only weakly? Maybe the spin of mu mesons is higher than one-half, but in such a case you have to worry about why the radiative cross sections are not large actually. Perhaps you would have to take into account certain internal structure of the mu meson. Whatever you may think, the problem of the mu meson is very challenging. There are also mysteries surrounding photons. If you look at the whole body of elementary particles including all the new particles, then the photon, which is most familiar to us, is in a sense the strangest of all. One of the mysteries of the photon is related to the concept of charge independence, which seems to work quite well in the cases of meson-nucleon and nucleon-nucleon interactions and also in a more comprehensive theory of particles. However, the introduction of electromagnetic interaction destroys the isotropy of isotopic spin space. This seems quite strange to me. I have no way of understanding this strange situation so far, and I cannot be very confident about the notion of isospin space, until there appears a good idea of explaining the peculiarity of electromagnetic interaction.

One of the mysterious properties of strongly interacting particles is that they can interact sometimes very strongly, and sometimes very weakly. I have been trying to give a possible interpretation of this remarkable difference in strong and weak interaction for the same kind of particles. Of course, there are a number of different approaches to this problem, but I take the way in which we suppose that the known principles and symmetry properties of present physics stay on, that is to say, special relativity and the principles of quantum mechanics stay on as far as they can, just to see how far we can go on this way. We consider only those particles which can interact strongly with each other. Each of them is supposed to have a certain kind of structure which is in conformity with special relativity. I know that this is not only a severe restriction, but also this gives rise to a very difficult problem connected with causality. I know very well how difficult it is to conform to causality in any nonlocal field theory. But, let us just try to see to what extent such a model will work.

Now, such a particle with internal structure could be expressed in terms of a wave function ψ, with a number of components depending on the coordinates of the particles as a whole and also on a number of internal variables. Of course, there may be extra degrees of freedom which have little to do with our ordinary space-time concept. For instance, isospin space might be something which cannot be incorporated into our notion of space-time in the narrow sense. For

the time being, at least, we have to deal with the isospin as such. However, we may suppose that there are other extra degrees of freedom which are more directly connected with our space-time world. In other words, it may be possible to single out those hidden variables which are called the internal coordinates in the narrow sense. Of course, we shall have to envisage eventually the over-all structure of the world underlying all the ordinary and extra degrees of freedom, but let us suppose for the moment that a subspace of the whole space underlying the extra degrees of freedom has the same structure as the external space-time world. If we adopt such an assumption, what we have to deal with is a kind of nonlocal field. However, we are not so much interested in the detailed procedure of formulation of such a theory as the new possibilities for understanding the behavior of particles in terms of nonlocal field theory. For the latter purpose, let us simplify the situation as much as possible. The field for each type of particle is a function ψ of not only one set of space-time coordinates X_μ, which is related to the motion of the particle as a whole, but also of variables ξ_μ which could be identified with internal coordinates designating a point in the internal world. The latter is supposed to have the same structure as the external space-time world. If we ignore all other extra degrees of freedom, the field can be denoted by $\psi(X_\mu,\xi_\mu)$ which may consist of a number of components.

Now let us introduce the following three drastically simplified assumptions:

(I) In the zero approximation, the field for each of the strongly interacting particles is written as a product of external and internal wave functions:

$$\psi(X_\mu,\xi_\mu)=\psi_{\text{ext}}(X_\mu)\psi_{\text{int}}(\xi_\mu).$$

Each of ψ_{ext} and ψ_{int} may have a number of components, which transform linearly under a Lorentz transformation. Thus, external spin and parity can be discriminated from internal as well as from total spin and parity by considering the Lorentz transformations of X_μ and ξ_μ separately as well as the simultaneous transformations of both.

(II) Each of the interaction terms in the Lagrangian or the Hamiltonian for a system of elementary particles should be constructed from three fields in such a way that it is a relativistically invariant product of two factors, of which one is an invariant product of three external wave functions and the other is an invariant product of corresponding internal wave functions in the case of strong interaction.

(III) In the case of weak interaction, the interaction term in the Lagrangian should again be relativistically invariant as a whole, but not necessarily be so for external and internal parts separately. For instance, it could be a scalar product of external and internal spinors, vectors etc. In the following, we shall consider only the case of scalar product of two spinor functions.

If, in addition to these three assumptions, we take into account the charge independence of pion-nucleon interaction, we obtain the following scheme as one of the simplest possible examples of assignment of external and internal wave functions to each type of strongly interacting particles:

	ψ_{ext}	ψ_{int}	total spin (and parity)
π	pseudoscalar	scalar	0^-
N	Dirac	scalar	$\frac{1}{2}$
Λ	scalar	Dirac	$\frac{1}{2}$
K	Dirac	anti-Dirac	$0^\pm$, $1^\pm$
⋮	⋮	⋮	⋮

Here anti-Dirac means that the wave function transforms like the charge conjugate of Dirac wave function. If we adopt this scheme, it follows from the second assumption that the strong interactions are restricted essentially to the cases such as $\bar{N}\pi N$, $\bar{\Lambda}\bar{K}N$, $\bar{K}\pi K\cdots$. Similarly, according to the third assumption, the weak interactions could be restricted to the cases $\bar{N}\pi\Lambda,\cdots$ and possiby $\pi\pi K$. All of these strong and weak interactions are acceptable and some of them, at least, are necessary in order to account qualitatively for both the production and decay processes of strongly interacting particles. The total spin and parity are conserved always, but the weak interaction is distinct from the strong interaction in that the former gives rise to the exchange of angular momentum of an amount $\hbar/2$ between external and internal degrees of freedom of interacting particles.

If we compare this scheme with the Gell-Mann-Nishijima scheme, we notice immediately that the strangeness quantum number is equal to twice the internal spin. As already mentioned, the isospin quantum numbers are left out of our scheme in view of the peculiarity of electromagnetic interaction.

Perhaps the most disappointing feature of our scheme is that we still cannot easily get rid of the well-known dilemma of theta and tau modes of decay of K mesons. I do not know whether a more complicated scheme involving higher spins could serve for the final solution of this crucial problem. We cannot yet tell whether we are already facing the problem which concerns the validity of the rigid framework of special relativity.

Pi Mesodynamics

F. E. Low
Massachusetts Institute of Technology, Cambridge, Massachusetts

I WOULD like to discuss in a qualitative way how far we can come in understanding the behavior of low-energy pi mesons. I would like what I say to be understandable to people who are not experts, and I hope that the experts present will forgive me for oversimplification and omission and even half-truths, though I hope there will not be very much of that.

The two real mesonic phenomena that take place at low energies are scattering of mesons by nucleons, and photoproduction of mesons. Clearly experiments with real mesons are a better way of trying to understand the meson-nucleon system than experiments involving virtual phenomena, such as the neutron-proton potential or the charge densities of neutrons and protons. These virtual phenomena are difficult to understand and interpret, and the connection between the numbers that one can measure and real behavior of mesons is not as clear. So I prefer to say very little about the virtual phenomena.

Let me give an example of the importance of real phenomena. By the study of meson-nucleon scattering in the last few years it has been possible to verify quite accurately an aspect of the theory originally proposed by Yukawa, which it had not been possible to verify by means of the study of virtual processes: In the language of field theory we now know that the meson-nucleon interaction is linear in the meson field; in particle language that the meson interacts by virtual absorption and re-emission. I will come back to this point very often. Almost all of what small success there has been in understanding the meson-nucleon system has been in terms of understanding this particular aspect of the interaction of mesons and nucleons.

Let me review briefly the experimentally known properties of pi mesons. The pi meson has spin zero. It has odd parity, which means that when a negative pion is absorbed by a proton, the result of this absorption being a neutron, the "orbital" parity of the system has to change. This has rather striking consequences for the interaction, because it means that a pi minus in an S state cannot be absorbed by a proton to become a neutron, because then the orbital parity cannot change. The pi minus can be absorbed from a P state, however, because then there is a negative orbital parity which can disappear in the absorption act. Therefore, if this picture of the meson-nucleon interaction is correct, one suspects immediately that P-wave effects will be important and S-wave effects less important. This is indeed the case; and I will come back to this again.

The meson has isotopic spin one, and the total isotopic spin of mesons and nucleons is conserved approximately in the course of their interactions. I would like to digress for a minute here and remind you that the conservation of isotopic spin is approximate. One hopes that the degree of approximation is measured by the fine structure constant, 1/137. There are known violations of isotopic spin conservation which are larger: the mass difference between neutral and positive charged mesons is about 3%. As may be seen simply from phase space arguments, this mass difference can induce violations of isotopic spin conservation of 9% in P-wave scattering. There is another, possibly large, violation of isotopic spin conservation which may turn up: there may be a difference of a few percent (how many no one knows) in the coupling of charged and neutral mesons to nucleons. Let me therefore interpose a warning in the interpretation of experimental data: details of the analysis of experiments that depend upon 5% accuracy are unreliable. For example, the determination of the isotopic spin $\frac{1}{2}$, P-wave phase shifts would be very hard even with considerably more accurate experiments, in view of the difficulties just mentioned. In this talk I will not be concerned with 5% effects, and, will assume that isotopic spin is conserved.

For general orientation as to orders of magnitude, Fig. 1 shows the total pi-plus and pi-minus cross sections on protons between zero and 1.8 Bev energy in the lab. Starting at the left: notice the very small cross sections at zero kinetic energy—already an indication that the S-wave scattering is anomalously small at low energies, and a strong hint of the dominance of the P-wave scattering. This is almost a direct verification from experiments of the Yukawa absorp-

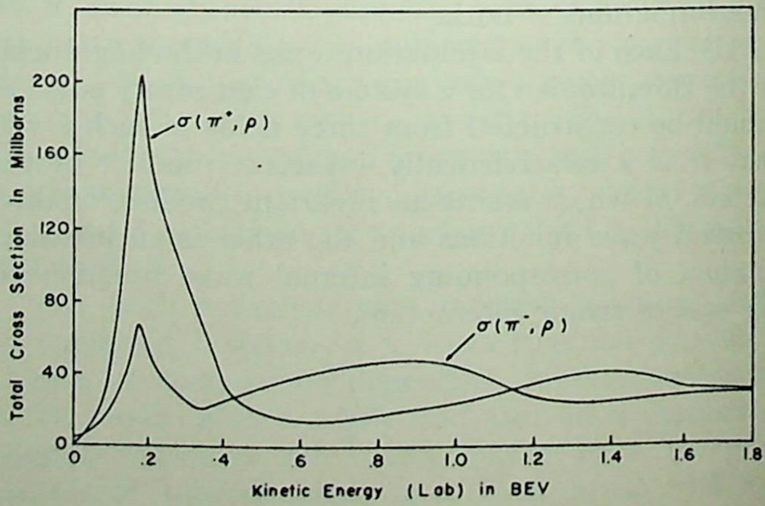


FIG. 1. Cross sections for scattering of π^+ and π^- mesons by protons as functions of meson kinetic energy in Bev in reference frame in which protons are initially at rest.

tion and re-emission mechanism mentioned at the beginning. Moving to the right in Fig. 1 to 200 Mev, the large maximum in the positive pi cross section, the famous 3–3 resonance, and its smaller reflection in the pi-minus cross section are seen. At still higher energies, above 400 Mev, the cross sections remain large. For example, at 800 Mev the pi-minus cross section is as large as it is at the resonance; this fact has some unfortunate significance for understanding of the low-energy phenomena. I return to this point at the end of my talk.

Very little of a specific nature is understood about the high energy region, however, so I will limit myself to energies smaller than 300 Mev. Low-energy pion-nucleon scattering can be understood extraordinarily well on the basis of the hypothesis first proposed by Brueckner, that the scattering goes predominantly through a P state of angular momentum $\frac{3}{2}$ and isotopic spin $\frac{3}{2}$, this state having the resonance shown in Fig. 1 at about 185 Mev in the laboratory. Now let us see just how well this appears to account for the data. Figure 2 shows three times the negative pi cross section and the positive pi cross section plotted against energy. Were the scattering all in the state of isotopic spin $\frac{3}{2}$ these curves would be identical. Their striking equality is evident. The experimental errors are not shown, but the curves overlap in the energy region under consideration when these are included.

In order to investigate the accuracy of the assumption of an angular momentum of $\frac{3}{2}$ we refer to Fig. 3, which shows the angular distribution in the positive meson scattering by protons. The experiment is at 189 Mev, which is very near the resonance. (This is somewhat of a cheat since the angular distribution will fit the $J=\frac{3}{2}$ assumption best near the resonance.) Were the scattering in a P state of angular momentum $\frac{3}{2}$, the angular distribution would have the form $1+3\cos^2\theta$. Again the agreement with experiment is extremely good. Figures 2 and 3 emphasize the fact that the Brueckner resonance hypothesis fits the data extremely well. One can really predict what will happen on the basis of this resonance. The fact is that the (3,3) state dominates the entire low energy picture,

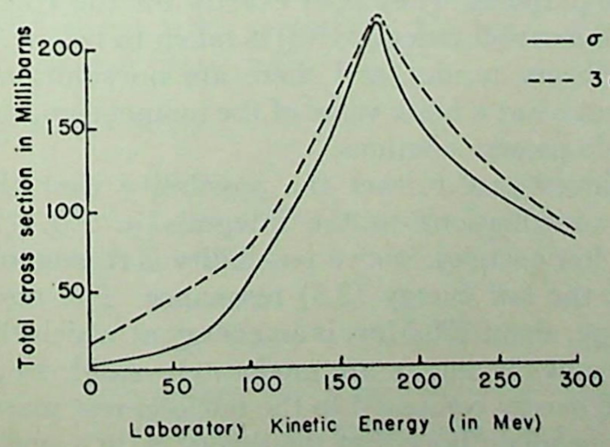


FIG. 2. Comparison of cross section for π^+ scattering by protons with three times that for scattering of π^- by protons in the low-energy resonance region, abscissas as in Fig. 1.

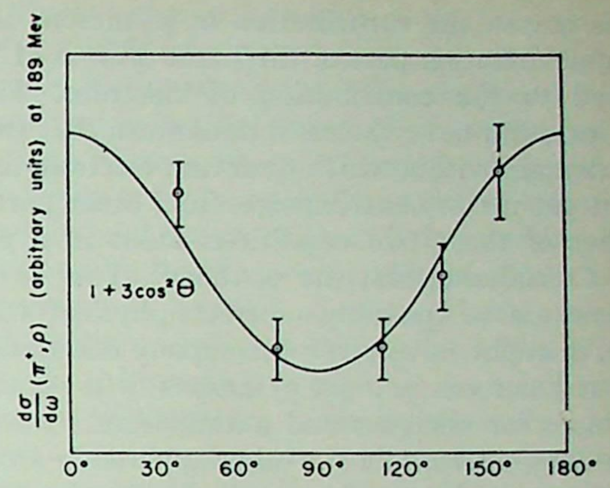


FIG. 3. Angular distribution of the scattering of π^+ mesons by protons as measured at a meson kinetic energy of 189 Mev.

including phenomena as far removed from meson scattering as the Compton scattering of photons by protons. Thus, if one can understand the resonance, to a great extent one also understands the low-energy scattering. We can understand a little bit about the resonance, and I would like to explain what that is.

Let me outline a group of questions, which represent a possible program for understanding moderate energy pion phenomena. First, can we predict that there will be a resonance in the (3,3) state, and can we calculate the meson energy at which this resonance will take place? Second, having predicted a resonance, can we calculate the shape of the resonance, for example, its width? Third, having taken care of the resonance and its shape in the problem of meson-nucleon scattering, can we calculate other processes involving real mesons at low energies, such as photomeson production from nucleons? Fourth, can we say anything at all about fine details of the pion-nucleon scattering? By fine details I mean, for example, S wave scattering, P-wave scattering in the nonresonant states, D-wave scattering, and so forth. These are the problems that a theory faces in trying to account for the low energy data. Unfortunately the theoretical situation is very complicated. Here is how things stand.

In the first place we are faced with the unfortunate situation of not knowing the correct theory that we have to use to try to explain the phenomena. Two theories have become fair game for theoretical physicists. The first of these is the relativistic, symmetric pseudoscalar meson theory, the γ_5 theory; the second is the Chew theory. Neither is a correct theory in the sense that quantum electrodynamics is a correct theory. Let me emphasize the two reasons that quantum electrodynamics is what we call a correct theory. In the first place the fine structure constant is small, 1/137, so we can calculate in a power series, or a quasi-power series, and get accurate answers. The second reason is that the mass ratio of the particles that are involved in quantum electrodynamics to the particles which we like to keep out of quantum electrodynamics (such as the pi meson) is small, about 1/273.

For this reason the contribution of pi mesons to the vacuum polarization part of the Lamb shift is $(1/273)^2$ compared to the contribution of electrons. This is simply luck, but nevertheless it does mean that there is a wide domain within which quantum electrodynamics does not get major contributions from other particles.

Neither of these two conditions holds in a meson theory. Considering first the γ_5 theory. This, as far as we know, is a possibly correct physical theory. That is, it might have given a complete description of reality. It has room in it for pi mesons; it is covariant; it has room for nucleons and antinucleons. Unfortunately, it does not contain the other apparently strongly interacting particles which we now know exist. Therefore, it is a model which does not describe the real world. It is a model which might have described things over a large energy range, but which in fact does not. The mass ratio of the electron to the pi meson, the relevant mass ratio of quantum electrodynamics, is 1/273. The mass ratio in the pion theory case is effectively about $\frac{1}{5}$; therefore, pi-meson phenomena probably may not be isolated from the effects of other particles.

The second trouble is that the pion-nucleon interaction is a strong one. The interaction constant, instead of being 1/137, is about 14. These difficulties do not mean that the γ_5 theory is a useless theory. If we can deduce consequences from it which do not depend very strongly on whether there are other particles which are strongly coupled, then these consequences are useful, and certainly much more believable than consequences which do depend on whether there are other particles. Because of the strong coupling, we cannot really calculate anything with this theory; all we can do is prove theorems and sum rules, such as the dispersion relations. Happily, everything we *can* do with the pseudoscalar theory does not depend manifestly on what particles really exist at higher energy. We conclude that the theory may be incorrect, or it may be only incomplete. It could be that an essentially correct theory could be made by adding curious particles and their interactions to the pion-nucleon interaction already included. This may or may not be the case.

The second theory which has become fair game in the last few years is one which has existed for some time, but was first taken up seriously by Chew four years ago. In this the nucleon is assumed to have an infinite mass; that is, there is no recoil in the theory, no nucleon, antinucleon pairs, and no curious particles. In order to calculate anything in this theory one must introduce a cutoff; otherwise, the theory is divergent. The importance of this theory is that Chew was able to obtain with it quite striking agreement with experiment. He was able to calculate the (3,3) phase shift, and obtain the resonance. Also, he was able to calculate photomeson production with good accuracy. There are two parameters in the theory, a coupling constant f^2 and a cut-off energy ω_m. It is really amazing that this theory works, because in calculating with it (by expanding in powers of the coupling constant), almost every integral diverges very strongly, and hence depends strongly on ω_m. Apparently, therefore, the results of the theory depend strongly on what really goes on at high energy, on recoil, on antinucleons, on things which are not adequately described by the theory and which violate the conditions of the theory. Yet agreement is obtained, and it is not completely obvious why this is the case. I will try to communicate to you some of my feelings about why this works.

The reason the Chew theory works is that there exists a representation in which the virtual integrations, because of the resonance, are to a good approximation kept in the low-energy region. In this representation the equations of the static theory become

$$\frac{e^{i\delta_\alpha}\sin\delta_\alpha}{q^3v^2(q)}=\frac{\lambda_\alpha}{\omega_q}+\frac{1}{4\pi^2}\int_1^\infty\frac{d\omega_p}{pv^2(p)}\times\left[\frac{\sigma_\alpha(\omega_p)}{\omega_p-\omega_q-i\epsilon}+\sum_\beta\frac{A_{\alpha\beta}\sigma_\beta(\omega_p)}{\omega_p+\omega_q}\right], \quad (1)$$

where δ_α are the phase shifts, ω_q and q the meson energy and momentum, and $v^2(q)$ is the cut-off function. The subscripts α and β refer to the four states of different total angular momentum and isotopic spin. The meson mass is taken to be unity. The constants λ_α and $A_{\alpha\beta}$ are known and are given by

$$\lambda_\alpha=\tfrac{2}{3}f^2\begin{pmatrix}-4\\-1\\-1\\2\end{pmatrix}\text{ for the }\begin{pmatrix}11\\13\\31\\33\end{pmatrix}\text{ states}$$

$$A_{\alpha\beta}=\tfrac{1}{9}\begin{pmatrix}1&-4&-4&16\\-2&-1&8&4\\-2&8&-1&4\\8&2&2&1\end{pmatrix}.$$

Finally, f^2 is the pseudo-vector coupling constant, $f^2=0.08$, and σ_α the total cross section in the state α.

The derivation of these equations is unimportant for present purposes. They hold exactly for the Chew theory. If the cut-off function $v^2(q)$ is taken to be one, a divergent theory results, and there are no solutions. If $v^2(q)$ is cut off at a finite value of the momentum, the equations do possess solutions.

Let us investigate further the possibility that the dominant contributions to the integrals in Eq. (1) come from low energies. Such a possibility is reasonable because of the low energy (3,3) resonance. The resonance energy, about 200 Mev, is an energy at which the assumptions of the theory are moderately valid, since it is a small energy compared to the nucleon rest mass; therefore, one might hope that the theory is, in a sense, self-contained at low energies. If the (3,3) resonance is assumed to dominate the integrals, Eq. (1) may be

written in a much simpler form,

$$\frac{e^{i\delta_{33}}\sin\delta_{33}}{q^3}\cong\frac{4}{3}\frac{f^2}{\omega_q}+\int_1^{\sim 3}\frac{d\omega_p}{p}\frac{\sigma_{33}(\omega_p)}{4\pi^2}\times\left[\frac{1}{\omega_p-\omega_q-i\epsilon}+\frac{1}{9}\frac{1}{\omega_p+\omega_q}\right]$$
$$+\text{high-energy contribution (h.e.c.)} \quad (2)$$

$$\frac{e^{i\delta_{31}}\sin\delta_{31}}{q^3}\cong\frac{e^{i\delta_{13}}\sin\delta_{13}}{q^3}\cong-\frac{2}{3}\frac{f^2}{\omega_q}+\frac{4}{9}\int_1^{\sim 3}\frac{d\omega_p}{p}\frac{\sigma_{33}(\omega_p)}{4\pi^2}\frac{1}{\omega_p+\omega_q}+\text{h.e.c.} \quad (3)$$

$$\frac{e^{i\delta_{11}}\sin\delta_{11}}{q^3}\cong-\frac{8}{3}\frac{f^2}{\omega_q}+\frac{16}{9}\int_1^{\sim 3}\frac{d\omega_p}{p}\frac{\sigma_{33}(\omega_p)}{4\pi^2}\cdot\frac{1}{\omega_p+\omega_q}+\text{h.e.c}, \quad (4)$$

where σ_{33} is the total cross section in the (3,3) state, and the meson energy corresponding to the momentum q is denoted by ω_q. The cut-off function has been omitted from these equations since it is sensible to take it as one in the low-energy region. Whether these equations are useful depends on the high-energy contribution. If the high-energy contribution is important, then the static theory has no hope of working. If the high-energy contribution is unimportant, then the static theory can work.

I have said that one can deduce these equations from the static theory, and that these equations do in a sense close in on themselves at low energies. It does not follow from this that the equations are correct. We would like to have a derivation of these equations which starts from more reasonable assumptions than those of the static theory. How can one hope to derive these equations? The only derivation that can be given, as far as I know, is based on the dispersion relations about which you have heard so much at this meeting. The dispersion relations have apparently been proved, so I will accept them as correct. Examples of dispersion relations are given in Eq (5).

$$T_j(\nu,\kappa^2)=B_j(\nu,\kappa^2)+\frac{1}{\pi}\int_{1-\kappa^2/M}^{\infty}d\nu' ImT_j(\nu',\kappa^2)\frac{(\nu'+\nu)\pm(\nu'-\nu)}{\nu'^2-\nu^2-i\epsilon}, \quad (5)$$

where

$$\kappa^2=\tfrac{1}{4}(\Delta p)^2 \quad \nu=\omega_L-\kappa^2/M.$$

Here ω_L represents the total laboratory energy (excluding the nucleon rest mass), and $(\Delta p)^2$ is the momentum transfer squared, in the center-of-mass system. The quantities T_j represent the four-scattering amplitudes which correspond roughly to linear combinations of the two isotopic spin states, $\frac{1}{2}$ and $\frac{3}{2}$, and to spin-flip and non-spin-flip scattering. The amplitudes B_j are the Born approximation values of the scattering amplitudes, which may be calculated from pseudoscalar meson theory. ϵ is the usual real positive infinitesimal.

These relations give the scattering amplitude as a function, not quite of energy and angle, but of energy and momentum transfer squared. There are various difficulties associated with these dispersion relations. One is that the integrals of Eqs. (7) may not converge. If they fail to converge, these dispersion relations are incorrect. In such a case dispersion relations which are presumably correct may be obtained if a convergence factor is placed under the integral sign, and the Born approximation amplitudes are modified by unknown functions of the momentum transfer. Such equations are shown in Eqs. (6).

$$T_j(\nu,\kappa^2)=B_j'(\nu,\kappa^2)+\frac{\nu^2}{\pi}\int_{1-\kappa^2/M}^{\infty}d\nu' ImT_j(\nu',\kappa^2)\frac{(\nu'+\nu)\pm(\nu'-\nu)}{(\nu'^2-\nu^2-i\epsilon)\nu'^2}, \quad (6)$$

where B_j' represent the modified Born approximation amplitudes.

The experimental situation suggests that for one of the four amplitudes, one needs to use the dispersion relation (6), but that for the other three, one may use the dispersion relations (5). These, incidentally, are not the dispersion relations originally published by Goldberger; those were for the forward direction, corresponding to $\kappa^2=0$. Here the momentum transfer is different from zero, and one must integrate the variable ν' down to essentially zero kinetic energy. Since, one cannot have a finite momentum transfer at zero kinetic energy, this implies that one must have some prescription for analytically continuing the amplitudes $T_j(\nu',\kappa^2)$ into a region where the momentum transfer is too large for the available energy. At present we are unable to continue these functions analytically into a very large unphysical region (though the recent derivations of the dispersion relations may reveal the correct rule for doing this). However, we may still obtain a useful set of equations by differentiating various numbers of times with respect to κ^2 at $\kappa^2=0$. In general, we can obtain by subtraction of such differentiated equations an equation in which the leading term on the left is the amplitude we wish to investigate. If the D waves and higher waves are small, for example, it is clear that by subtraction we can obtain equations for the S-wave amplitudes and P-wave amplitudes at low energies. On the right side of these equations will be integrals over imaginary parts of these amplitudes. These imaginary parts can be calculated by unitarity; they are related to weighted sums of partial total cross sections. A "partial total cross section" means a cross section in a state of definite total angular momentum;

it is total in the sense that it includes inelastic as well as elastic effects.

The S- and P-wave dispersion equations are quite complicated and will not be listed here. However, again one might hope that in the integrals that occur on the right sides of these equations (if the imaginary parts of the amplitudes are replaced by cross sections), the 3–3 resonance is dominant when the energy ν' is not too high. If this hope is correct, one can calculate the right-hand sides of the equations in a simple way by dropping everything except terms referring to the 3–3 state. By this method one again obtains equations for which the dominant contributions come from low energies, and one can again make nonrelativistic approximations. If one assumes the nucleon mass to be large compared to the pion momenta, one obtains exactly the equations obtained before from the static theory, Eqs. (2), (3), and (4).

I would like to emphasize the difference between this and the previous derivation. Previously, we neglected terms of order $1/m$ in the pseudoscalar theory and obtained a static theory. Such a process is really self-contradictory, however, because the virtual integrations which must be carried out in order to solve the equations involve the high-energy region in which the initial assumptions are invalid. The dispersion relation method is a different thing; here we have neglected $1/m$ terms only with the assumption that a real observed resonance makes this neglect valid. This deduction of the static theory is on a quite different basis from the usual one. I hope that that point is clear. Although I believe it is a reasonable deduction as far as it goes, it is not a complete deduction, since it depends, in addition to the neglect of high-energy integrals, on the existence of a low-energy resonance.

As we shall see, the location of this resonance is very sensitive to high energies; put differently, Eq. (2) is very insensitive to the location of the resonance.

The above argument shows roughly why the Chew theory agrees with experiment. I realize that it is an involved and complicated argument; I apologize for the complications. Of course, the assumptions made here may not be right. The high-energy contributions we have neglected in the dispersion relations are different from those neglected in the Chew theory. The equations which were deduced from the Chew theory contain high-energy contributions from the states (3–3), (3–1), (1–3), and (1–1). The dispersion relations we have deduced contain high-energy contributions from whatever happens in the world, however, not just from what happens in a static P-wave theory. In order to compute these high-energy contributions one must integrate over sums of partial cross sections. These partial cross sections will be multiplied by functions of the orbital angular momentum l; for example, we might have a term of the following form:

$$(l+1)(l+2)(l+3)(l+4)\sigma_{l+1}. \quad (7)$$

If we can show that the contributions from such terms are indeed small, then we will have understood to some extent why the static theory is not inconsistent with experiment. I will come back to this at the end of my talk.

Now let us return to the discussion of what may be calculated in the γ_5 and Chew theories. The first thing we try is to predict the resonance. In the γ_5 theory this cannot be done. Calculations which produce a resonance may be made by keeping only certain terms; one way is the Tamm-Dancoff method, which does produce a resonance in the (3,3) state. I don't want to say anything against this method, because it is an extremely good one for calculating when the coupling is weak. When the coupling is strong, however, nobody knows how to calculate anything in relativistic field theories. It has been shown by Wyld that the Tamm-Dancoff calculation of the resonance in pseudoscalar meson theory is invalid. Wyld calculated the complete fourth order of the pseudoscalar theory and found that the contribution of the resonance producing terms are canceled almost exactly by terms which are neglected in the ordinary Tamm-Dancoff calculations. Therefore the enhancement which had been produced in the (3–3) state in the γ_5 theory disappeared. When we can't show that the resonance exists, we can't calculate the resonance energy either.

In the Chew theory the situation is somewhat different. In this static theory there are three or four quite independent methods of calculating. None of these is numerically reliable, but all of them agree that there is a resonance, and all of them can make use of the same coupling constant and the same cut-off energy to locate the resonance at approximately the right point. The different methods of calculating with the Chew theory give formulas for the resonance energy ω_r of the type

$$\omega_r \approx 1/f^2\omega_m,$$

where the energies are in units of the meson mass. This equation is not exact; different methods of calculation give results which differ by a factor of two or so. This is not a fundamental calculation of the resonance energy. Assuming the coupling constant to be fixed by some other method (in practice, this may be done by photoproduction measurements), one can always obtain the correct resonance energy by choosing ω_{max} appropriately. However, the theory does give a resonance in almost every limit. That is the main difference between the static theory and the γ_5 theory. One might tentatively conclude that if the γ_5 theory could make itself look like the Chew theory it would have a chance, but nobody really knows now how to make it do that.

I hope I've made it plain that one can't calculate the resonance energy in either theory. However, if one is given the resonance energy from experiment, what can be done then? Then on the basis of the assumptions

that I have outlined, one can hope that the integrals in Eqs. (2), (3), and (4) exhaust the complete dispersion integrals, that is to say, that the high-energy contributions to these equations are indeed negligible. One can then hope that the low-energy behavior of T_{33} can be calculated from Eq. (2), and that is, in fact, the case. The shape of the resonance can be calculated from this equation. There are many simplifications in Eq. (2). For example, as ω approaches zero, the behavior of the left-hand side is given by the term, $(4/3)\ (f^2/\omega)$. In other words, if the scattering amplitude is considered as an analytic function of the complex variable ω, then the singularity of the function at $\omega=0$ is known. Furthermore, the resonance is a narrow resonance, the width being given by the small coupling constant, f^2. (The resonance is at low energy not because of the size of f^2, but because $\omega_{\max}$ is large.) Thus, there is a resonance with a narrow width, and if a narrow width is assumed in Eq. (2), it follows that the resonance has the shape given by the following "effective range" formula

$$\frac{4}{3}f^2 \cot\delta_{33}\frac{q^3}{\omega}=1-\omega/\omega_r. \tag{8}$$

This equation may be written in the form

$$\frac{4}{3}q^3/\omega \cot\delta_{33}=\frac{1}{f^2}(1-\omega/\omega_r). \tag{9}$$

One sees that the left side of Eq. (9) must be a linear function of the energy. This equation determines two things. First, it gives the shape of the resonance, and second, it provides a method of determining the coupling from the intercept of the straight-line curve at zero energy. Figure 4 shows the experimental points plotted against energy. The plot is a very good straight line and its intercept gives you the coupling constant. (This coupling constant is known as the Watson-Lepore coupling constant.) The fact that it is $q^3/\omega \cot\delta$ which is a straight-line function of ω and whose intercept at $\omega=0$ determines f^2 rather than $q^3 \cot\delta$ (as in the nucleon-nucleon case) is an experimental verification of the linear interaction mechanism which I have already mentioned; it arises from the term λ_α/ω in Eq. (1), which, in turn, comes from those Feynman diagrams which directly represent the emission and absorption of the real meson by the nucleon. The actual plot is made against ω^* rather than ω for reasons which are not essential to this conclusion. Here $\omega^*=\omega+q^2/2m$ is the center-of-mass energy.

I cannot say very much about photomeson production because of time. This is unfortunate because the calculation of photomeson production is a real success of the present approach. Everything that was said about the scattering can be translated to the photomeson production problem; on the basis of this type of formulation one may calculate photomeson production to about 15% accuracy.

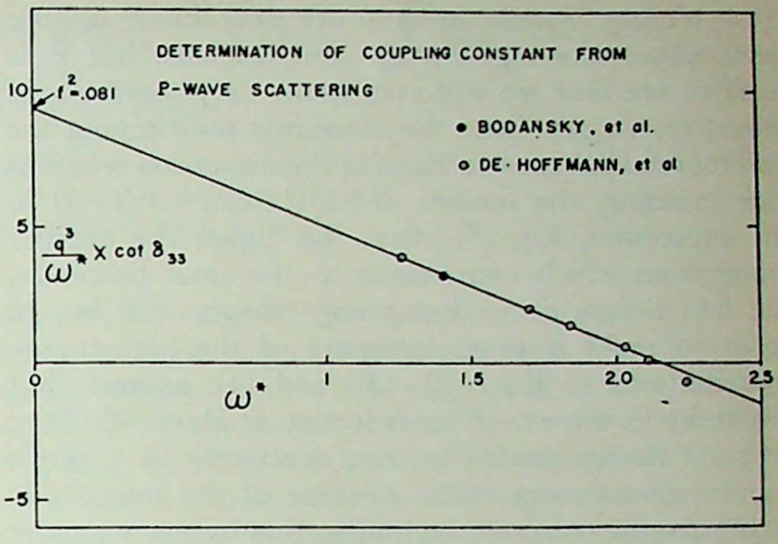


FIG. 4. Graph based on Eq. (9) showing determination of the Watson-Lepore coupling constant from experimental data.

Now let me discuss the question of the fine details. How accurate is the whole approach? How far can we go with it? Can we say anything about S waves at all? Again, on the basis of theory, the honest answer to almost all these questions has to be no, with perhaps one very interesting exception. On the basis of the dispersion relations plus certain assumptions about the high frequency behavior of scattering amplitudes, one can calculate one S-wave number. (This has been shown by Goldberger.) One can calculate apparently the threshold value of the charge exchange scattering amplitude. This amplitude may be written as an energy integral over the difference of two cross sections. Such a determination is successful because the integral converges at 2 Bev, whereas the cross sections are still large at 2 Bev. There is no evidence that they are not essentially of the same order of magnitude at 200 Bev, and there is no reason to have expected any assumption about asymptotic forms to have been useful at so low an energy as 2 Bev. There are plenty of particles still to be made; there are plenty of strong couplings left. Therefore, we cannot be sure the integral receives no large contributions from unexplored high-energy regions. Nonetheless, the integral apparently converges and gives a value of the charge exchange scattering length in agreement with that given by Orear. This calculation, I would say, is an isolated S-wave success; it simply happened to work. If you like, you can say this is an indication that the pseudoscalar theory is not quite as crazy as it might have been because this S-wave calculation involves the same coupling constant as the P-wave calculations and essentially the same singularity [the term involving f^2/ω in Eq. (1)] in the scattering amplitude.

Let us turn now to the last question: How accurate is this whole approach? That question is very hard to answer, because in order to answer it one must know the partial cross sections for high angular momenta (D waves, F. waves, etc.) at energies above 400 Mev; whereas all that is measured in this region is the total

cross section. Sooner or later we may know a little more about the high-energy contributions, but it is hard to see how we will ever know very much more. However, we see from the constants multiplying the different partial cross sections in the dispersion relations [for example, the factors, $(l+1)(l+2)(l+3)(l+4)$ in the expression, Eq. (7)] that the higher the angular momentum which contributes to the total processes, the less accurate our low-energy theory will be. In order to make a rough estimate of the high-energy contributions to Eqs. (2), (3), and (4), assume that the hump in the $\pi^- - P$ cross section at about 900 Mev (Fig. 1) results primarily from scattering in a single angular momentum state. Because of the magnitude of the rise in the cross section in this hump, we must assume that the angular momentum involved is not too low; a $D_{\frac{5}{2}}$ wave is about the lowest we can assume. Assuming then the hump to be caused by $D_{\frac{5}{2}}$ scattering, what is the effect of this h.e.c. on the low-energy amplitudes? One finds that the order of magnitude of these high-energy contributions is not very small. The contribution of the $D_{\frac{5}{2}}$ hump to Eqs. (2), (3) and (4) is, roughly, $f^2/3$. In the equation for the (3,3) amplitude, this is not too serious, since the two important terms on the right of Eq. (2), $4/3f^2/\omega$ and

$$(4\pi^2)^{-1}\int d\omega_p \sigma_{33}(\omega_p)/_p(\omega_p-\omega)$$

are both positive at low energy and add to a value of roughly eight or more times $f^2/3$. However, in the equations for δ_{13}, δ_{31}, and δ_{11} the high-energy contribution is more serious, for in these equations the f^2/ω terms are negative, and, when combined with the integral term, give results of the order of $f^2/3$.

This result is unfortunate. What it says is that the predictions that can be made on the basis of the static theory about the small phases are quite unreliable. It says that precise predictions about the large (3,3) phase also cannot be made. It had been hoped that for the P waves some miracle would happen so that the low-energy description really was self-contained. This miracle has only happened with low accuracy. Because of the (3,3) resonance, as I have tried to make clear, we can obtain a rough understanding of low-energy scattering and photoproduction. To do better, to make precise predictions, to calculate S waves, I believe we will have to understand much more about what really goes on at high energy.

Weak Interactions between "Old" Particles and Beta Decay

LOUIS MICHEL

Institute of Physics, Lille, France

DR. Yang will speak to us on the "new" particles. In this talk, I shall speak only of the "old" particles. To simplify the matter, the main part of this talk will be concerned with the question whether it is possible to explain all the properties of the "old" particles by adding only *one* interaction to the gravitational, electromagnetic, and nuclear forces. This interaction, which I shall call the "Fermi interaction," was first proposed by Fermi in 1934, and represents the most successful application of the ideas of the quantum theory of fields outside of quantum electrodynamics. In this paper (Fer 34),* Fermi uses a plain second quantization language; for instance, he does not speak about negative energy states; we shall follow here his example. For shorthand, I shall call fermion a spin $\frac{1}{2}$ particle. I shall follow Fermi's procedure in trying to construct a relativistic interaction between the four fermions in the β-decay process,

$$n\rightarrow p^{+}+e^{-}+\nu,$$

for example. The way to proceed is to take two fermion wave functions, say, those of the heavy particles, and to construct with them the five covariants: a scalar S, a vector V, a skew-symmetrical tensor T, an axial vector or pseudovector A, and a pseudoscalar P. Now take the corresponding covariants constructed with the light particle wave functions, and contract them with the first set of covariants. In this way, one obtains five scalars and five pseudoscalars; which are scalars and which are pseudoscalars is really a matter of convention, because to define parity you have to define it relative to something. In the literature, one set, the so-called "even coupling" has been used much more frequently than the "odd coupling," but *a priori* one set is as good as another. As a matter of fact Fermi just happened to choose an odd coupling, the coupling with the vector current of the nucleon field (in analogy with electrodynamics) and he left the other couplings to be worked out by the reader. The main conclusions of Fermi's paper were the following: when the initial and final nuclei have the same spin and parity, the electron spectrum is essentially the statistical spectrum. More precisely

$$P(E)dE=\frac{1}{2\pi^3}|M|^2g_V{}^2\mathfrak{F}(Z,E)E(E^2-1)^{\frac{1}{2}}\times[(W-E)^2-\nu^2]^{\frac{1}{2}}\left(1-\frac{\nu}{E(W-E)}\right),\quad(1)$$

where $\hbar=c=m_e=1$; the coupling constants are g_S, g_V, g_T, g_A, g_P; ν is the neutrino mass; W is the total energy released by the nucleus; $\mathfrak{F}(Z,E)$ is the influence of the Coulomb interaction due to the charge Ze of the nucleus, i.e., $\mathfrak{F}(0,E)=1$; and $|M|^2$ depends on the nucleus and is called the nuclear matrix element.

The constant g appears to be very small: 6×10^{-12} in this unit system or 3×10^{-49} erg cm^3 in cgs.

A precise determination of the electron spectrum at the end point would give a measurement of the neutrino mass, and also, if $\nu\neq0$, it would impose a choice between even and odd couplings. Indeed, as was pointed out by Konopinski and Uhlenbeck (Kon 35), one must change ν into $-\nu$ in (1), to obtain the corresponding formula for even coupling. But it is *not* true that a precise measurement of the end point of the electron spectrum would enable one to decide between the two following possibilities for a neutrino theory: the Dirac theory (used by Fermi) in which there are two kinds of neutrinos, corresponding to particle-neutrinos and antiparticle-neutrinos, and the Majorana theory (Maj 37, and add to it Rac 37) in which all neutrinos are identical.

It happens that ν is very small, maybe zero ($\nu=m_\nu/m_e=10^{-3}$). We shall neglect it from now on.

At the time when the Fermi theory was proposed, the experimental data were very sketchy, and it was not clear that the theory worked well. However, it soon had two successes: it allowed for β^+ decay which was discovered a few months later, and it predicted K capture, which was discovered three years later. Since it has survived many difficulties connected with imprecise experimental data; these difficulties are of historical interest only, and will not be mentioned here.

As a first approximation in his calculations, Fermi noted that the velocity of the nucleon in the nucleus is nonrelativistic, and its effect either averages out or can be neglected; further, he noted that the wavelengths of the outgoing electron and neutrino are of the order of 100 times larger than the nuclear radius, and therefore he kept only the constant term in the expansion of the wave functions of the light particles. When these approximations are made and a nonvanishing decay rate is obtained, the transition is called *allowed*. It requires for the nucleus the following selection rules:

g_S and g_V couplings: $\Delta I=0$, *no* parity change (Fermi selection rules,

* A complete bibliography is listed at the end of the article.

g_T and g_A couplings: $\Delta I=0$ or 1, $0\to 0$ transitions excluded, *no* parity change (Gamow-Teller selection rules, Gam 36),

g_P does not induce allowed transitions.

Transitions not allowed are called first, second, $\cdots n$th forbidden according to the order of the first nonvanishing term. An important development of the beta-radioactivity theory was the study of the forbidden transitions.

But here we shall not study chronologically the evolution of Fermi theory. Instead, we shall try to answer the very important question: which of the five possible Fermi couplings exists in nature and with what relative magnitude? This question has provoked many papers. Here, to answer it, we shall give the most straightforward proofs, without caring for historical ones.

First, what do the allowed transitions tell us on this subject?

There are many allowed transitions (e.g. He^6, $\Delta I=1$) which obey Gamow-Teller and not Fermi selection rules. Since 1949 several allowed transitions, $0\to 0$, have been found: they do not obey Gamow-Teller selection rules; e.g,

$$O^{14}\to\beta^+ + N^{14} \quad \text{(She 49)}$$

$$C^{10}\to\beta^+ + B^{10} \quad \text{(She 52)}$$

$$Cl^{34}\to\beta^+ + S^{34} \quad \text{(Arb 53)}$$

$$Al^{26}\to\beta^+ + Mg^{26} \quad \text{(Kav 55).}$$

Therefore, the two following hypotheses are *ruled out*:

$$g_S=g_V=0 \quad \text{and} \quad g_T=g_A=0.$$

In allowed transitions, there are interference between g_S and g_V couplings and also between g_T and g_A couplings. What is the experimental value of those interference terms (called Fierzterms, Fie 37). To be precise, here is the most general electron energy spectrum for allowed transitions:

$$P\pm(E)dE = 2\pi^{-3}(X\mp X'/E)\mathfrak{F}(Z,E)E(E^2-1)^{\frac{1}{2}}(W-E)^2dE, \quad (2)$$

where $\pm$ indicates β^+ or β^- decay and

$$X=(g_S^2+g_V^2)|M|^2+(g_T^2+g_A^2)|\mathbf{N}|^2$$

$$X'=2(1-\alpha^2Z^2)^{\frac{1}{2}}(g_Sg_V|M|^2+g_Tg_A|\mathbf{N}|^2)$$

in which M and $\mathbf{N}$ stand for these nuclear matrix elements connecting initial and final states,

$$M=\int\psi_f{}^*\psi_i d\tau \quad \text{and} \quad \mathbf{N}=\int\psi_f{}^*\sigma\psi_i d\tau. \qquad (2')$$

The most sensitive way to measure g_Tg_A is to compare the rate of competing β^+ decay and K capture for a $\Delta I=1$, no, transition (no g_S and g_V contribution). The interference terms are of opposite signs in the two cases and the nuclear matrix elements are the same. This study has been done first by Sherr and Miller (She 54—see also Kre 54, All 55). Their result is that one g is 0.00 ± 0.02 the other. (Results from the spectrum shape of $Cu^{64}\beta\pm$ and P^{32} by Mah 52, Dav 53, Poh 56, are a little less precise.)

It is more difficult to measure g_Sg_V (Mah 52 on N^{13} and S^{35}, when both Fermi and Gamow-Teller couplings compete, say that $g_Sg_V/(g_S^2+g_V^2)$ is "substantially less than one"). It will require more experimental work on $0\to 0$ transitions, which were discovered quite recently and have short lifetimes. The precise shape for these transitions is not yet known, and we have to get what information we can from the activity, i.e. the ft value. The integration of the spectrum [Eq. (2)] yields for a $0\to 0$ transition,

$$2\pi^3\log_e 2=ft[g_S^2+g_V^2+2(1-\alpha^2Z^2)^{\frac{1}{2}}g_Sg_V\langle E^{-1}\rangle]|M|^2, \quad (3)$$

where

$$f=\int_1^w \mathfrak{F}(Z,E)E(E^2-1)^{\frac{1}{2}}(W-E)^2dE,$$

$$\langle E^{-1}\rangle=f^{-1}\int_1^w \mathfrak{F}(Z,E)(E^2-1)^{\frac{1}{2}}(W-E)^2dE,$$

and t is the half-life. Thus a plot of $2\pi^3\log_e 2/ft|M|^2$ as a function of $(1-\alpha^2Z^2)^{\frac{1}{2}}\langle E^{-1}\rangle$ for different ($0\to 0$, *no*) transitions, will give the desired ratio $2g_Sg_V/(g_S^2+g_V^2)$ as indicated in Fig. 1. The quantities Z, f, t, $\langle 1/E\rangle$ are experimental quantities and $|M|^2$ can be computed to good accuracy, usually a few percent. An analysis of Gerhart and Sherr (Ger 56) using the ft values for O^{14}, Al^{26}, and Cl^{34} yields for the ratio 0.00 ± 0.15.

Therefore, we need essentially only one g from each pair g_S, g_V and g_T, g_A. This leaves us with four possibilities, and we must find out which of these possibilities fits with experimental data.

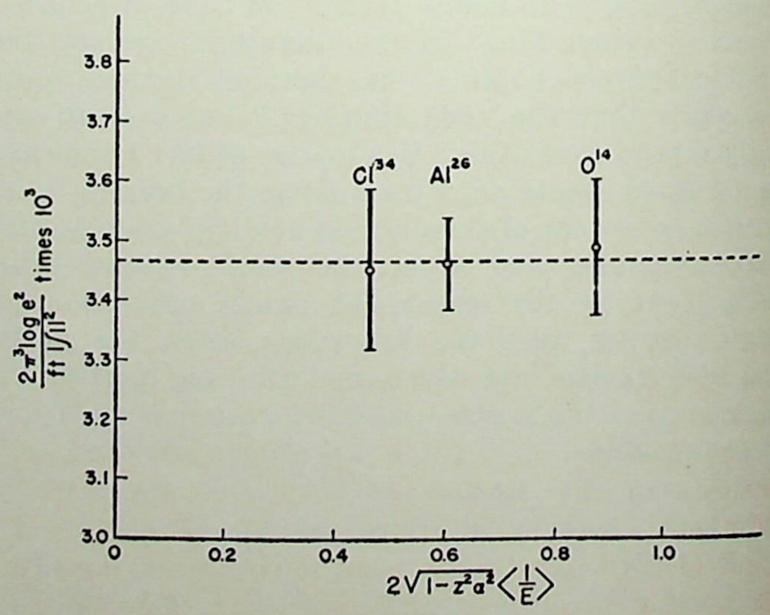


FIG. 1. Up-to-date discussion of Gerhart and Sherr, Bull. Am. Phys. Soc. Ser. II, 1, 195 (1956). The slope of the dotted line gives the cross term g_Sg_V. The interaction with the vertical axis gives $g_S^2+g_V^2$ in the unit system $\hbar=m_e=1$ and the second.

The most direct answer has been given, after years of effort by experimental physicists, by detecting the recoil of the nucleus in beta decay. The experimental data on recoil is most conveniently expressed in terms of a parameter α which appears in the correlation function $f(\theta)$ between the electron and neutrino directions of emission. For allowed transitions

$$f(\theta)=1+\alpha v_e \cos\theta \qquad (4)$$

where v_e is the electron velocity and

$$\alpha=[(g_V^2-g_S^2)|M|^2+\tfrac{1}{3}(g_T^2-g_A^2)|\mathbf{N}|^2](X\mp X'/E)^{-1}$$

in which X and X' are the same as in (2).

The results of Rustad and Ruby (Rus 53, 55) and of Allen and Jentschke (All 53) for He^6 (a Gamow-Teller transition) unambiguously point in favor of g_T. Using this result one may apply this analysis to nuclei in which the transition obeys both Fermi and Gamow-Teller selection rules. The experiments on the recoil of Ne^{19} by Alford and Hamilton (Alf 54) and by Maxon, Allen, and Jentschke (Max 55) as well as the correlation studies of Robson (Rob 55) in the neutron decay unambiguously lead to $g_S \gg g_V$, so that we may say that the allowed transitions are essentially induced by g_T and g_S.

To find the ratio of these coupling constants we shall neglect g_V/g_S and g_A/g_T. The integration of Eq. (2) then yields

$$A=ft[|M|^2+R|\mathbf{N}|^2]=2\pi^3 g_S^{-2}\log_e 2. \qquad (5)$$

Here A is independent of the decaying nucleus, and $R=(g_T/g_S)^2$. The value of ft is given by experiment, and the nuclear matrix elements have to be computed theoretically. Now in a plot of A *vs* R, each nucleus is represented by a line. If our assertions are correct, all these lines should pass through a single point which defines a unique value of A and R. The data (Fig. 2) show indeed that the lines do converge. The lines correspond to the neutron decay, N^{15} and F^{17} decays (mirror transition of one closed-shell $\pm$ one nucleon). Those nuclei are the most reliable for calculation of $|\mathbf{N}|^2$. The dotted line represents He^6, the horizontal lines O^{14}, Al^{26}, Cl^{34}. From Fig. 1, we can conclude qualitatively that g_S and g_T are not very different and $|g_S|<|g_T|$. Similar discussions have been done since 1950 by Feingold and Wigner (Fei 50), Moszkowski (Mos 51), Trigg (Tri 52), Kofoed-Hansen and Winther (Kof 52, 53, 56, Win 52), Nataf and Bouchez (Nat 52), Blatt (Bla 53), Wu (Wu 54), Feenberg *et al.* (Fee 55, Bol 55). Here we have followed Gerhardt's (Ger 54) notation.

The most recent, and not yet published, systematic analysis to my knowledge is Kofoed-Hansen and Winther's (Kof 56. In their notation $B=A/(1+R)$, $x=R/(1+R)$). They use more mirror transitions and their value of $|\mathbf{N}|^2$ is semiempirically deduced from the measured magnetic moment of the nuclei, according to the Bohr Mottelson collective model. They claim that an analysis with $g_V=0$ is not consistent with quoted experimental errors. This is not surprising since physicists often overestimate the accuracy of their experimental or theoretical results! However, after a more systematic discussion, not neglecting g_V/g_S, these authors obtain a better fit of the data† with the introduction of an admixture of g_V interaction with $g_V/g_S=0.15$. This result needs to be confirmed. I am not competent to discuss the assigned value of experimental errors, but I can give an example of overestimation of accuracy by theoretical physicists. The nuclear matrix elements $|M|^2$ for $0\to 0$ transition between mirror level of nuclei is easy to compute, with a good accuracy and its value is actually independent of nuclear models (Rad 53, M Do 54). However, for g_S interaction, the accurate expression for m is

$$M=\int \psi_f^*\beta\psi_i d\tau$$

to which the previously given expression (2′) is only an approximation, used in all the papers under review. The theoretical value of the accurate expression does depend on the nuclear model. Professor Jensen tells me that Stech has made an unpublished computation for O^{14} which shows a 4% difference between the accurate and the approximate expressions for M.

Let us summarize what we learn from the allowed transitions.

The Fermi interaction for β decay is mainly in g_S and g_T. It is possible that $g_V=g_A=0$. The tentative limits are

$$g_V/g_S=0.00\pm 0.15, \quad g_A/g_T=0.00\pm 0.02,$$
$$(g_S^2+g_V^2)^{\frac{1}{2}}/|g_T|=0.90\pm 0.04.$$

In order to learn about the sign of g_S/g_T and the value of g_P we have to turn to the forbidden transitions, which I did not discuss, but the study of which nicely confirms all of the above conclusions. Some authors (e.g., Pea 53, Tio 55) say that the sign of g_S/g_T can be different for β^+ and β^- processes. They use a language different from that used here and by the majority of physicists. Both conventions agree for β^- decay. In chronological order, Morita, Fujita, and Yamada (Mor 53) using shell model theory, *j-j* coupling, claimed that they could fit the spectra of Fe^{59}, Rb^{87}, Te^{99}, and Cs^{137} with a mixture of g_S and g_T only and a negative sign for their ratio. Radium E, a nickname for Bi^{210} has, because of its queerly shaped spectrum excited a lot of interest, but with its spin having been

† The data they use are for O^{14}, "older" than that I have used in Fig. 1. This may modify the numbers to be given in the published paper. I thank these authors for communication of their manuscript and discussion. I thank J. B. Gehrart from whom I learned the new datum for O^{14}: unpublished data on the $C^{12}(He^3,n)O^{14}$ threshold by D. A. Bromley which gives for O^{14}, $ft=3091\pm 100$ sec. In the vernacular of physicists working in β decay, ft is expressed in the unit system $\hbar=m_e=1$, and second for unit of time.

measured to be unity (in 1954), a number of papers based on the wrong spin assignment became obsolete. Recently Lee-Whiting (Lee 55) fitted its spectrum with $g_S/g_T>0$, and criticized the conclusions of Morita *et al.* on the grounds that the nuclear model which they used may not be very suitable for the nuclei they considered. Fujita (Fuj 55) re-examined the problem using the Bohr-Mottelson nuclear model and is stronger in his assertion that the sign is negative, while Takebe *et al.* (Tak 55) looked at RaE again, and concluded that it is impossible to determine the sign of g_S/g_T from this nucleus. As you see from the above, one may at least conclude that the prediction of the value and even the sign of the nuclear matrix elements for the forbidden spectra requires a much better knowledge of nuclear forces inside the nuclei than we now possess (see also Mor 56). The same problem arises when one tries to say something about g_P, the pseudoscalar coupling. Of course g_P cannot be very much larger than the other g's; otherwise the allowed transition spectra would not have the shape they have. However if one wants to set an upper limit to $|g_P/g_T|$ one has to go to pretty high values. For example, the neutron decay spectrum which has the allowed shape sets an upper limit of the order of 100. Looking at more complicated nuclei, the value depends quite sensitively on the type of assumption made about the nuclear forces: many papers have stressed this point. Another possibility is to compare systematically the transitions $\Delta I=0$, *yes* (where g_p can compete) with other first forbidden transitions (where g_p plays no role). Some people see a difference between the two classes, others do not. Furthermore one still has to prove that the difference is due to the presence of g_P. I think it is fair to conclude that the question of the value of g_P is still completely open, and strictly speaking, there is no compelling evidence that $g_P\neq 0$.

To finish with the question of the nature of the Fermi coupling, we have to say if it is an "even" or "odd" coupling. As we have seen, if the mass of the neutrino were not zero, it would be possible to answer this question just from the shape of the allowed spectrum, mainly near the end point. However, the data are compatible with a vanishing neutrino mass. Several methods have been studied for the determination of the neutrino mass: closed cycles of nuclear reactions and decays; *ft* values and spectrum shape of low energy allowed transition. The best case is H^3. It must be emphasized that the influence of the neutrino mass on *ft* can be exactly compensated by a Fierz interference term. The most sensitive method is therefore to study the shape of the spectrum. The most recent study (Ham 53, see also Lan 52) gives $\nu=m_\gamma/m_e<0.0003$ for odd couplings, $\nu<0.0010$ for even couplings. If the neutrino has a zero rest mass, one will have to dare to measure the polarization of the emitted neutrino before deciding between even and odd couplings. It becomes a fairly academic problem. However, the fashion nowadays is to consider the nonconservation of parity for weak couplings. In Fermi theory of beta radioactivity, this means that we take a mixture of even and odd couplings. The situation is then different and requires study of the polarization of any of the four involved particles. Lee and Yang (Lee 56) have proposed several experiments for testing this hypothesis. Before leaving the topic of beta radioactivity, I must mention that before 1947 the Fermi theory ran second-best to the Yukawa theory of beta radioactivity

$$(n,p)\rightarrow Y\rightarrow(e,\nu)$$

because the latter not only explained this process, but also the decay of the then known meson. Discovery of the π meson, and its pseudoscalar character, showed that the Yukawa theory could not explain beta decay, and this brought the "phenomenological" Fermi theory back. There are many reasons for this (e.g. Cai 51a). An extension of Furry's theorem (Mic 52) requires a specific covariant nature for the Y meson, in order to obtain what was discussed by g_S or g_T interaction. But it might still be possible to explain all β-decay data by couplings through virtual, hypothetical mesons. Some Japanese physicists have galantly supported this possibility (Ino 48, Nak 50, Ume 52, Tan 48, 52, 53a, 53b, Oga 56). It was soon strengthened by the observation of Klein (Kle 48), Puppi (Pup 48, 49), Tiomno and Wheeler (Tio 49), Lee, Rosenbluth, and Yang (Lee 49) that identical Fermi couplings could explain μ-meson physics.

By 1947 it was shown that μ^- mesons, when stopped in matter, are captured by the nucleus from a Bohr orbit and a noninteracting energetic light particle is emitted, the final nucleus getting rid of its excitation by the evaporation of a few nucleons. This is quite similar to the process of K capture. The literature on μ capture is quite abundant, and I shall just mention the most straightforward comparison made of the coupling strengths in an experiment by Godfrey (God 53). The experiment is essentially this

$$\mu^-+C^{12}\rightarrow B^{12}+\nu$$
$$B^{12}\rightarrow C^{12}+e^-+\nu.$$

The decay rate of B^{12} was known and Godfrey had to measure the capture rate of μ capture when the final state of the nucleus is the ground state of B^{12}. (What he measured actually, was the transition rate to bound states of B^{12}; the two rates must be practically equal.) Since the nuclear matrix elements are the same in the transitions (11) and (12) (after a small correction due to the very different wavelengths of the neutrinos in those two reactions), the ratio of the transition rates gives directly the ratio of the strength of the coupling producing them. Since $\Delta I=1$, *no*, it is only g_T and Godfrey's result is

$$\frac{g_T^2 \text{ for } \beta \text{ radioactivity}}{g_T^2 \text{ for } \mu\text{-meson capture}}=0.94\pm 0.26.$$

Therefore this experiment is well explained by the hypothesis that the "same" Fermi coupling exists between $npe\nu$ and $np\mu\nu$. This shows a puzzling kinship between the μ meson and the electron.

There is no contradiction between present μ-capture experimental data and theory. It seems not possible in the near future to make experiments leading to an unambiguous choice of the five couplings constants for μ capture. But some crucial experiments can be done to see if the "same" choice of Fermi coupling constants as that of β decay can explain μ capture. This is thoroughly discussed in a forthcoming paper of Primakoff (Pri 56). Among the best experiments are the capture of μ^- mesons in hydrogen and deuterium.

The Fermi interaction is also quite successful in explaining the μ decay

$$\mu^{\pm} \rightarrow e^{\pm} + \nu + \nu.$$

To a good approximation it predicts the following electron spectrum (Mic 49)

$$P(E)dE = 4(E^2/W^4)[3(W-E)+\tfrac{2}{3}\rho(4E-3W)]$$

where the parameter ρ depends on the nature of the coupling. This shape forms a one-parameter family of curves, and is shown in Fig. 3. As two neutrinos are emitted, one must distinguish two cases: (i) the two neutrinos are identical, i.e., both are "neutrinos" or both are "antineutrinos" or the neutrino is a Majorana particle, in which case it turns out that $0 \leq \rho \leq \frac{3}{4}$, and (ii) the neutrinos are not identical in which case $0 \leq \rho \leq 1$. The experimental spectra have the shape given above, and the nearly two dozen experiments measuring ρ roughly agree with the most recent and precise published results of Sar 55, Cro 55 which are, respectively, 0.64 ± 0.10 and 0.50 ± 0.10.

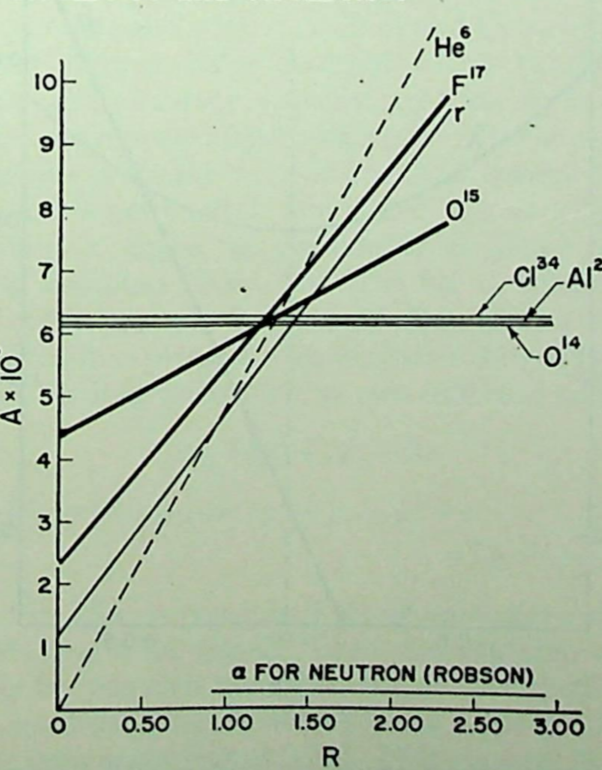


FIG. 2. In Gerhart's notation (Ger 54), plot of R *versus* A for some super-allowed transitions. (The curve marked "r" should be marked "n.")

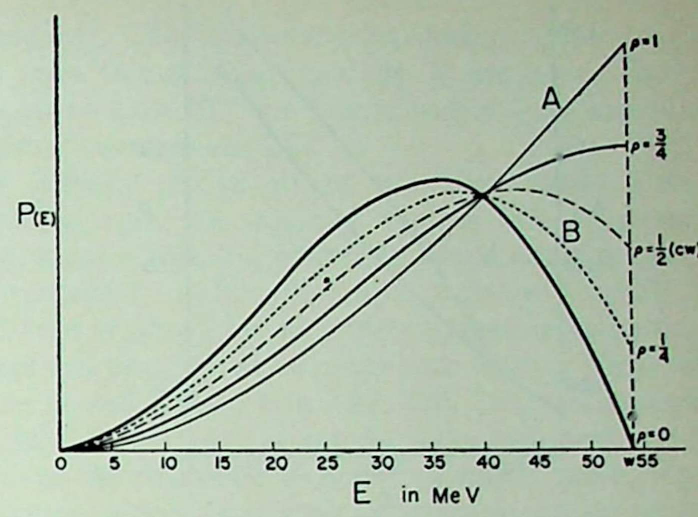


FIG. 3. Electron spectrum from μ-meson decay.

If one wants to compare the Fermi interaction in ($npe\nu$) and ($\mu e\nu\nu$), one must recognize the following complication: the interaction is sensitive to the order in which the particles are written in the interaction term. On changing the order (from a very general theorem of group theory, see also Fie 37) the new five invariants will be linear combinations of the old five invariants. (Everybody in β radioactivity respects a traditional order.) Therefore, to make a comparison between different Fermi interactions, one must adopt a one-to-one correspondence between the sets of four particles one wants to study. There are $4! = 24$ such correspondence. However, only three classes of correspondences give different physical results, and therefore to make a comparison, one must state the class chosen, and decide whether the neutrinos emitted in the decay are identical or not.

Finally a test of the equality of the Fermi interaction strengths in ($npe\nu$) and ($\mu e\nu\nu$) is the experimental value of the parameter λ (Mic 52), which represents essentially the ratio of the μ-meson decay ft value to that of the neutron decay.

$$\lambda = \frac{2^9(ft)_n\tau'_n}{\mu^5\tau_\mu \log_e 2} = \frac{2^3 A}{(1+3R)\mu^5\tau_\mu \log_e 2},$$

where the parameters A and R have been defined before. Experimentally, $\lambda = 1.05 \pm 0.14$ from the neutron data and 1.11 ± 0.06 from superallowed β decays. The recent data used is $\mu = 206.9 \pm 0.2$ (Bar 56); τ_μ is the μ-meson mean life; and for the neutron half-life $\tau'_n = 732 \pm 90$ sec (Spi 55).

It turns out to be possible to explain all of the data in terms of the same interaction, provided one makes the correspondence between pairs as follows:

$$(np)-(e\nu)-(\mu\nu).$$

(Triangle of interaction of Tiomno and Wheeler, Tio 49.)

In fact there are two possible fits. In the first one, in

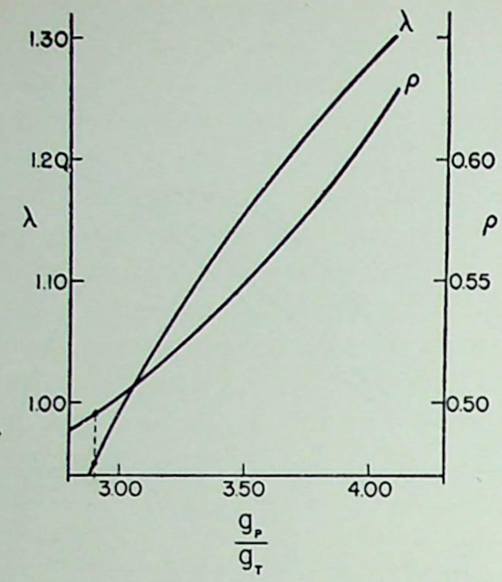


FIG. 4. Identical neutrinos. $g_S/g_T=0.9$. Correspondence $\begin{pmatrix} n & p & e & \nu \\ \mu & \nu & e & \nu \end{pmatrix}$.

which the neutrinos in the μ decay are identical, to agree with the experimental value of ρ and λ we need $|g_P/g_T|$ of the order of 3. In the second case, when the neutrinos are taken to be distinguishable, one can obtain a fit by assuming essentially no pseudoscalar coupling.

The situation is summarized in Figs. 4 and 5. It has not essentially changed since the summary of the same analysis I published with A. Wightman two years ago (Mic 54). I apologize for keeping your attention on such old stuff while so many new things are discovered every day in the field of "new" particles. Is it not a striking coincidence that the "same" Fermi coupling is able to explain all data up to date on β radioactivity and μ-meson physics? However, there is one process that has challenged theoretical physicists: the π-meson decay; mainly the fact that the ratio of the experimental values of the rates (Lok 55):

$$\frac{\text{Rate }(\pi\rightarrow e+\nu)}{\text{Rate }(\pi\rightarrow\mu+\nu)} \quad\text{or}\quad \frac{\text{Rate }(\pi\rightarrow e+\nu+\gamma)}{\text{Rate }(\pi\rightarrow\mu+\nu)} \leqslant 5\times 10^{-5}$$

cannot be accounted for within the framework of the universal Fermi interaction as is described here.

So presently physicists are confused. Before the discovery of new particles, the hypothesis of Fermi interaction looked so promising that some physicists had become more ambitious: they tried to predict in an *a priori* way between which sets of four fermions the universal Fermi interaction exists. (For instance we know from experiments that $\mu+p\rightarrow p+e$; $\mu+n\rightarrow n+e$, $\mu\rightarrow 3e$ do not exist.) Such attempts have been either in the direction of adjusting phases under space or time reversal (Yan 50, Gam 50, Cai 51b, 52) (this is of course equivalent to saying (Wic 52) arbitrarily that there exist super-selection rules forbidding non-occurring processes) or by adding a conservation law similar to the conservation of nuclear and electric charge (Mah 52, Zel 53). These attempts emphasize the existence of selection rules and it seems fair to say that we do not understand them.

Theoretical physicists have also attacked the problem of justifying the form of the Fermi coupling, by a variety of proposals involving all kinds of symmetry principles, some simple, some very sophisticated. The oldest and simplest was proposed in 1941 by Critchfield and Wigner (Cri 41):

$$g_S=-g_A=-g_P, \quad g_V=g_T=0.$$

All papers proposing a coupling with $g_S\neq 0$, $g_T\neq 0$, have arrived essentially at the same conclusion (up to a sign) $|g_S|=|g_T|=|g_P|$; $g_V=g_A=0$ (Pur 51, 52, Mah 52, Pry 52, Pea 53, Fin 53, Ste 55, Tio 55, Pea 55). As we have seen $|g_S/g_T|$ is actually somewhat <1, and g_P/g_T is experimentally unknown. The calculated value of ρ is $\frac{3}{4}$, which is barely inside the experimental error, and the calculated value of λ is 4/3, which appears to be ruled out by experiment. Then some physicists have said more: corrections for radiative mesic effects are important for β radioactivity and not for μ-meson decay. These corrections were actually calculated (Fin 54, Ger 55, Ste 56, Ros 57) in the hope that they would raise $|g_S/g_T|$ from 0.9 to 1.0. Although the results depend on the meson theory used, the meson corrections are larger than expected and they seem to give $|g_S/g_T|>1$.

As far as fitting the new particles into the framework of the Fermi interaction is concerned, I will not say much. The lifetimes for the decays of the new particles seem to agree well with this kind of coupling, but

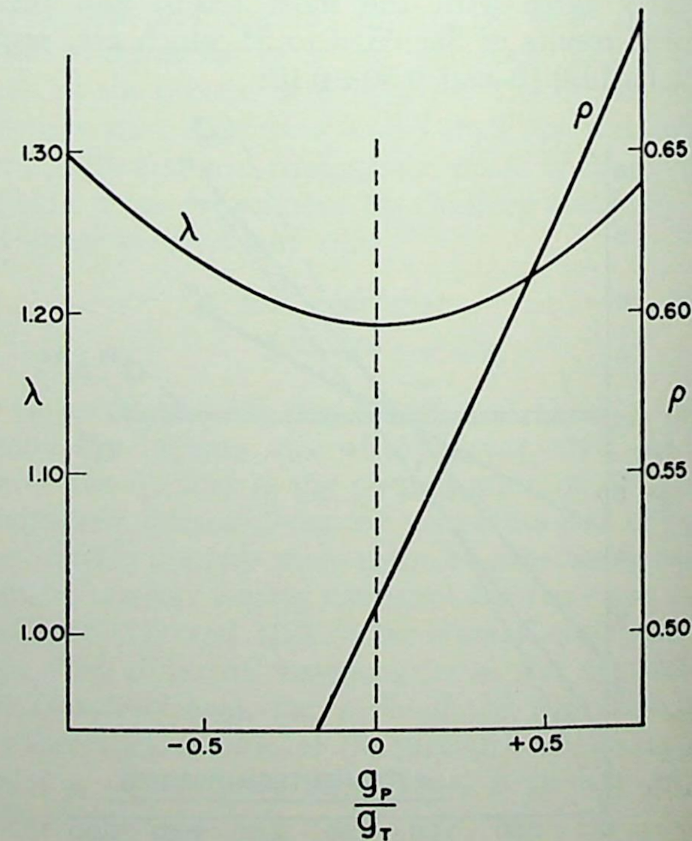


FIG. 5. Neutrinos indistinguishable. $g_S/g_T=-0.9$. Correspondence $\begin{pmatrix} n & p & e & \nu \\ \mu & \nu & e & \nu \end{pmatrix}$.

fitting them in does not suppress the obstacle that is caused by the experimental result that the rate for decay $\pi\rightarrow e+\nu$ or $\pi\rightarrow e+\nu+\gamma$ is roughly 5×10^{-5} times smaller than the rate $\pi\rightarrow\mu+\nu$. (By playing with interactions you can explain this, and some of the lifetimes, but then you have to forget about explaining beta radioactivity, which is no advantage!)

In conclusion, an analysis through a Fermi coupling, shows that the rates of decay processes lead to a striking equality for the strengths of the weak couplings which are responsible for them.

Since we have no consistent quantum theory of fields interacting through a Fermi coupling it seems that we do not understand what could be an universal Fermi interaction (the electric charge e is truly universal and we know that e is the renormalized coupling constant). We can even say: since the analysis through a Fermi coupling uses only the first Born approximation of an unrenormalizable coupling, it is mainly phenomenological and has not much to do with quantum field theory.

To conclude, it seems to me more suitable to speak about the neutrino and the μ meson since they do not belong to the subject of other reports at this conference. Indeed, with the electron, they are the only known particles with no strong coupling interactions: this, in some respects, makes easier theoretical interpretation of their experimental properties! They are not often taken care of in general schemes on strange particles. As example of exception, see Schwinger, this conference and Sachs (Sac 55) which suggests giving them a half-integer strangeness.

This year everybody has heard about the neutrino, since it made front-page news in the press! It has become a tame particle and with it, Reines and Cowan (Rei 56) from Los Alamos have changed a few protons into neutrons and electrons (inverse reaction to β decay). We know even more about neutrinos. After seven years of controversy on the rate of double β decay, it seems now that experimental physicists agree and their conclusion is there is no double β decay without neutrino emission (Aws 56). Double β decay is the second order process: 2 neutrons (bound in a nucleus) are changed into protons with emission of light particles. This can be done according to two different schemes

$$2n\rightarrow 2p^{+}+2e^{-}+2\nu \qquad (5a)$$

$$n+n\rightarrow p^{+}+e^{-}+\nu_M+n\rightarrow p^{+}+e^{-}+p^{+}+e^{-} = 2p^{+}+2e^{-}. \qquad (5b)$$

Reaction (5b) is possible if all neutrinos are identical (ν_M= Majorana neutrino). Then the neutrino emitted virtually by the first neutron can be absorbed by the second neutron. This does not seem to occur. Double β decay goes according to (5a). *This requires that there are neutrinos and antineutrinos in nature* (indeed a neutrino emitted by a neutron cannot be reabsorbed by another neutron; only its antiparticle can be absorbed). This is also the conclusion of Davies (Dav 56) who found experimentally a much too low (0.9 instead of 2.6×10^{-45} cm^2) cross section for pile neutrinos (from β^- decay) on Cl^{37}.

I already spoke about neutrino mass. For the neutrino spin we can say that a direct interaction n, p, e, ν, for spin $>\frac{1}{2}$ neutrions would not give the right energy spectrum for allowed beta decay (Ono 51).

There is also a lot of μ-meson physics going on and there will be much more in the near future. The μ-meson mass is well known now (Bar 56). Its spin assignment is old (Chr 41) but it can be soon measured again as well as its magnetic moment (μ-mesic atom, electromagnetic μ pair production). Electromagnetic radiative corrections can be studied experimentally and theoretically (μ-mesic atom, μ decay). The radiative correction to μ-meson decay has been computed by (Abr 51, Len 53, Beh 56). In the last paper it is proved that for a given coupling, these corrections can change ρ up to 4%, and that the study of the low energy part of the spectrum would give more information on the coupling. I have already spoken about the proposed μ-meson capture experiment. The μ meson, although indeed a very strange particle (what is its kinship with the electron?) is becoming also a tame particle, and will probably help us to probe the nuclei.

Note added in proof.—Since this has been written, there has been important news. Parity nonconservation has been discovered.

BIBLIOGRAPHY

Fer 34—E. Fermi, Z. Physik **88**, 161 or Nuovo cimento **11**, 1.
Kon 35—E. J. Konopinski and G. E. Uhlenbeck, Phys. Rev. **48**, 7.
Gam 36—G. Gamow and E. Teller, Phys. Rev. **49**, 895.
Fie 37—M. Fierz, Z. Physik **104**, 553.
Maj 37—E. Majorana, Nuovo cimento **14**, 171.
Rac 37—G. Racah, Nuovo cimento **14**, 322.
Chr 41—R. F. Christy and S. Kusaka, Phys. Rev. **59**, 414.
Cri 41—L. Critchfield and E. Wigner, Phys. Rev. **60**, 412.
Kle 48—O. Klein, Nature **161**, 897.
Ino 48—T. Inoue and S. Ogawa, Progr. Theoret. Phys. (Japan) **3**, 319.
Kon 48—
Pup 48—G. Puppi, Nuovo cimento **5**, 587.
Tan 48—Y. Tanikawa, Progr. Theoret. Phys. (Japan) **3**, 338.
Fei 49—A. M. Feingold and E. P. Wigner (mimeographed notes).
Fin 49—R. Finkelstein and M. Ruderman, Phys. Rev. **75**, 1958 (1949).
Lee 49—Lee, Rosenbluth, and Yang, Phys. Rev. **75**, 905 (1949).
Mic 49—L. Michel, Nature **163**, 959.
Pup 49—G. Puppi, Nuovo cimento **6**, 194.
She 49—Sherr, Muether, and White, Phys. Rev. **75**, 282 (1949).
Tio 49—J. Tiomno, Phys. Rev. **76**, 856 (1949).
Gam 50—A. Gamba, Nuovo cimento **7**, 919.
Nak 50—Nakamura, Fukuda, Ono, Sasaki, and Taketani, Progr. Theoret. Phys. (Japan) **5**, 740.
Yan 50—C. N. Yang and J. Tiomno, Phys. Rev. **79**, 495 (1950).
Abr 51—A. Abragam and J. Horowitz, J. phys. radium **12**, 952.
Cai 51a—E. Caianiello, Phys. Rev. **81**, 625 (1951).
Cai 51b—E. Caianiello, Nuovo cimento **8**, 634.
Mos 51—S. A. Moszkowski, Phys. Rev. **82**, 118 (1951).
Pur 51—D. L. Pursey, Phil. Mag. **42**, 1193.
Ono 51—Ono, Progr Theoret. Phys. (Japan) **6**, 238.
Cai 52—E. Caianiello, Nuovo cimento **9**, 336.

Kof 52—O. Kofoed-Hansen and A. Winther, Phys. Rev. **86**, 42 (1952).
Lan 52—L. M. Langer and R. J. P. Moffat, Phys. Rev. **88**, 689 (1952).
Mah 52—H. M. Mahmoud and E. J. Konopinski, Phys. Rev. **88**, 1266 (1952).
Mic 52—L. Michel, *Progress in Cosmic Ray Physics* (North Holland Publishing Company, Amsterdam), Chap. III.
Nat 52—R. Nataf and R. Bouchez, Compt. rend. **234**, 86; Phys. Rev. **87**, 155 (1952).
Pry 52—M. H. L. Pryce, Z. Physik **133**, 309.
Pur 52—D. L. Pursey, Physica **18**, 1017.
She 52—R. Sherr and J. Gerhart, Phys. Rev. **86**, 619.
Tan 52—Y. Tanikawa, Progr. Theoret. Phys. (Japan) **10**, 362.
Tri 52—G. L. Trigg, Phys. Rev. **86**, 506 (1952).
Ume 52—H. Umerawa, Progr. Theoret. Phys. (Japan) **7**, 551.
Wic 52—Wick, Wightman and Wigner, Phys. Rev. **88**, 101 (1952).
Win 52—A. Winther, Physica **18**, 1079.
All 53—J. S. Allen and N. K. Jentschke, Phys. Rev. **89**, 902 (1953).
Arb 53—W. Arber and P. Stähelin, Helv. Phys. Acta **26**, 433.
Bla 53—J. M. Blatt, Phys. Rev. **89**, 83 (1953).
Dav 53—J. P. Davison and D. C. Peaslee, Phys. Rev. **91**, 1232 (1953).
God 53—T. N. K. Godfrey, Phys. Rev. **92**, 512(L) (1953).
Fin 53—R. Finkelstein and P. Kavs, Phys. Rev. **92**, 1316 (1953).
Ham 53—Hamilton, Alford, and Gross, Phys. Rev. **92**, 1521 (1953).
Kof 53—O. Kofoed-Hansen and A. Winther, Dan. Mat. Fys. Medd. **27**, No. 14.
Len 53—A. Lenard, Phys. Rev. **90**, 968 (1953).
Mor 53—Morita, Fujita, and Yamada, Progr. Theoret. Phys. (Japan) **10**, 630.
Pea 53—D. C. Peaslee, Phys. Rev. **91**, 1447 (1953).
Rad 53—L. A. Radicati, Proc. Phys. Soc. (London) **A66**, 139.
Rud 53—A. Rudik, Doklady Akad. Nauk. (S.S.S.R.) **92**, 739.
Rus 53—B. M. Rustad and S. L. Ruby, Phys. Rev. **82**, 557 (1951).
Tan 53a—S. Tanaka and M. Itö, Progr. Theoret. Phys. (Japan) **9**, 169.
Tan 53b—Y. Tanikawa and K. Saeki, Progr. Theoret. Phys. (Japan) **10**, 232.
Zel 53—Ya. B. Zeldovich, Doklady Akad. Nauk. (S.S.S.R.) **91**, 1317.
Alf 54—W. P. Alford and P. R. Hamilton, Phys. Rev. **95**, 1351 (1954).
Fin 54—R. J. Finkelstein and S. A. Moszkowski, Phys. Rev. **95**, 1695 (1954).
Ger 54—J. B. Gerhart, Phys. Rev. **95**, 288 (1954).
Kre 54—W. E. Kreger, Phys. Rev. **96**, 1154 (1954).
Mdo 54—M. W. McDonald, thesis, Princeton.
Mic 54—L. Michel and A. S. Wightman, Phys. Rev. **93**, 354 (1954).
She 54—R. Sherr and R. H. Miller, Phys. Rev. **93**, 1076 (1954).
Wu 54—C. S. Wu, *Proceedings of the 1954 Glasgow Conference on Nucleon and Meson Physics* (Pergamon Press, London), Bellamy edition.
All 55—Allen, Burchman, Chachert, Munday, and Reasbeck, Proc. Phys. Soc. (London) **A68**, 681.
Bol 55—M. Bolsterli and E. Feenberg, Phys. Rev. **97**, 736 (1955).
Cro 55—Crowe, Helm, and Taufest, Phys. Rev. **99**, 872.
Fuj 55—J. Fujita, Progr. Theoret. Phys. (Japan) **13**, 260.
Ger 55—S. S. Gershtein and Ya. B. Zeldovich, Zhur. Exptl. Theoret. Phys. **29**, 698.
Iwa 55—Iwata, Ogamas, Okonogi, Sakita, and Oneda, Progr. Theoret. Phys. (Japan) **13**, 19.
Kav 55—Kavanagh, Mills, and Sherr, Phys. Rev. **97**, 248 (1955).
Lee 55—G. E. Lee-Whiting, Phys. Rev. **97**, 463 (1955).
Lok 55—S. Lokanathan and J. Steinberger, Phys. Rev. **98**, 240 (1955).
Max 55—Maxon, Allen, and Jentsche, Phys. Rev. **97**, 109 (1955).
Pea 55—D. C. Peaslee, Z. Physik **141**, 399.
Rob 55—J. M. Robson, Phys. Rev. **100**, 933(L) (1955).
Rus 55—B. M. Rustad and S. L. Ruby, Phys. Rev. **97**, 991 (1955).
Sac 55—R. G. Sachs, Phys. Rev. **99**, 1573 (1955).
Sar 55—Sargent, Rinehart, Lederman, and Rogers, Phys. Rev. **99**, 885 (1955).
Spi 55—Spirac, Sosnowsky and Prokofiw, Geneva Conference A, Conf. **8**, 650.
Ste 55—B. Stech and J. H. D. Jensen, Z. Physik **141**, 175.
Tak 55—Takebe, Nakamura, and Taketani, Progr. Theoret. Phys. (Japan) **14**, 317.
Tio 55—J. Tiomno, Nuovo cimento **1**, 226.
Tre 55—S. B. Treiman and H. N. Wyld, Jr., Phys. Rev. **99**, 1573 (1955).
Aws 56—M. Awschalon, Phys. Rev. **101**, 1041 (1956).
Bar 56—Barkas, Birnbaum, and Smith, Phys. Rev. **101**, 778 (1956).
Beh 56—Behrends, Finkelstein, and Sirlin, Phys. Rev. **101**, 866 (1956).
Dav 56—R. Davies, Sr., Bull. Am. Phys. Soc. Ser. II, **1**, 219 (1956).
Ger 56—J. B. Gerhart and R. Sherr, Bull. Am. Phys. Soc. Ser II, **1**, 195 (1956). Corrected for new data.
Kof 56—O. Kofoed-Hansen and A. Winther, Dan. Mat. Fys. Medd. **27**, No. 14.
Lee 56—T. D. Lee and C. N. Yang, Phys. Rev, **104**, 254.
Mor 56—M. Morita, Progr. Theoret. Phys. (Japan) **15**, 445.
Oga 56—S. Ogawa, Progr. Theoret. Phys. (Japan) **15**, 487.
Poh 56—Pohm, Waddell, and Jensen, Phys. Rev. **101**, 1315 (1956).
Pri 56—H. Primakoff (to be published).
Rei 56—F. Reines and C. L. Cowan, Nature **178**, 446.
Ros 56—M. Ross, Phys. Rev. **104**, 1736.
Ste 56—B. Stech, Z. Physik **145**, 319.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 2 APRIL, 1957

Present Knowledge About the New Particles

C. N. YANG

Institute for Advanced Study, Princeton, New Jersey

THIS discussion is divided into the following four sections: First an introduction to the subject, dealing mostly with generalities; then, a brief description of what seems at present to be established knowledge concerning the strange particles; thirdly, some topics currently under discussion by theoretical physicists; lastly, a question which has in the past year occupied a great deal of attention, namely, the identity of the K particles.

The discovery of the strange particles dates back to 1947, when Rochester and Butler[1] found, in a magnetic cloud chamber, V-shaped events in a penetrating shower induced by cosmic rays. This was followed in 1949 by the observation and identification[2] by Powell's group in England of the decay of the τ meson in electron sensitive emulsions. Experimental work on these strange particles, and others found later, rapidly expanded in the last few years. With the completion in 1953 of the cosmotron and in 1955 of the bevatron, experiments entered into a new era in which one can work with controlled beams of strange particles and in which one is able to make relatively accurate measurements on their interactions and decay. The scope and intensity of the experimental activities on this subject may be gauged from the fact that at the Brookhaven National Laboratory in the United States, 60% of the cosmotron time at present is devoted to the study of strange particles, the remaining time being divided between pion physics, radio chemistry, machine design, and other subjects.

What have we learned about these particles? We know the identity of quite a number of them; that is to say their masses, their charges, their lifetimes and decay modes. We know something about the processes by which they are produced. We could say that we know enough, so that there has emerged a common language among us in talking about them. We know enough to be able to plan controlled experiments which will tell us more about them. But the sum total of our knowledge goes very little beyond this. We are still confused by the identity of some of them, and we know very little about their spins and their parities. Above all, we do not have any idea or even any hint about the reason for their existence. Their relationship to each other and to the familiar particles do not seem to emerge in a simple pattern. Unlike Yukawa's π mesons, which fit into the physicist's picture of the world, the strange particles seem to be uncalled for, and unwanted. Do they presage the physicist's future understanding of small distances? Of the mass spectrum of particles? Or do they presage the breakdown of the concepts and laws that we have learned in the physics of nonstrange particles? These questions remain completely unanswered.

In view of our profound ignorance of what the whole subject is about, we must admit that we are very fortunate indeed to find in the experimental data many regularities, and many irregularities, which we hope may provide keys to a future understanding. We cherish such a hope mainly because of two facts: (a) The experiments consistently confirm a concept which we have had before, namely, that there is a natural classification of interactions into strong and weak interactions with very widely different orders of magnitude in their strengths. One senses that perhaps herein lies a very fundamental order of nature. (b) We recall that in the past, the concept of symmetry has played a very important role in physics. To quote a few examples: The theories of special and general relativity are founded upon such concepts of symmetry. Again, we could say that in essence the structure of the periodic table and therefore the properties of matter are largely determined by the isotropy of space, or in other words, by the symmetry of nature under space rotations. The prediction of the existence of the positron and of the antiproton were based upon the symmetry of nature under Lorentz transformations. Now it is observed that symmetries, and the related concepts of quantum numbers and selection rules are very much involved in the regularities and irregularities in the behavior of the new particles. We may therefore hope that the study of them will lead to a further understanding of nature.

To go on to the second topic, let us briefly outline our knowledge about the new particles. The well-established law of conservation of the number of nucleons requires that all particles have a "heavy particle quantum number" N. The nucleons have $N=1$. The pions, photons, electrons, neutrinos, and μ mesons have $N=0$. The antinucleons have $N=-1$. It is customary to call the strange particles with $N=1$ hyperons, those with $N=0$ mesons. Table I lists the experimental knowledge about the hyperons. Table II lists that about the positive and neutral mesons. Negative mesons are expected to behave like the charge conjugates of the positive ones, and they seem to do just that. They are therefore omitted from the table. In these tables, I have given numbers in round figures. More complete and detailed information with

[1] G. D. Rochester and C. C. Butler, Nature 160, 855 (1947).
[2] Brown *et al.* Nature 163, 82 (1949).

TABLE I. Experimental knowledge about hyperons.

Hyperon	Decay product	Q	Mean life
Λ^0	$p+\pi^-$	37 Mev	3.7×10^{-10} sec
Σ^+	$n+\pi^+$	110	$\sim1\ \times10^{-10}$
	$p+\pi^0$		
Σ^0	$\Lambda^0+\gamma$		short
Σ^-	$n+\pi^-$	110	$\sim2\ \times10^{-10}$
Ξ^-	$\Lambda^0+\pi^-$	~67	$\sim1\ \times10^{-10}$

references can be found in the various review compilations.[3]

In addition to tabulating the strange particles, we must say something about their interactions. There seem to be three kinds of interactions, the strong interactions, the electromagnetic interactions and the weak interactions. By the weak interactions we mean those responsible for the decay of the strange particles, for beta decay, for μ-meson decay, for π-meson decay, and for the μ-meson nucleon interaction. The reason why one feels some confidence in such a classification is that if one makes a rough dimensional analysis of the coupling strengths, one finds that the π-meson interactions and the production processes for the strange particles have intensities of around 1 to 10^{-1}; the electromagnetic interaction, as is well known, has strength 10^{-2}; and the weak interactions have intensities ranging from 10^{-12} to 10^{-14}. The wide gap separating the weak interactions from the other two forms such a striking and consistent pattern that one cannot but feel that there is some fundamental reason for it. A characteristic feature of the strange particles is that their production seems to fall into the class of strong interactions and their decays into the class of weak interactions. The practical manifestation of the weakness of the decay interaction is the fact that the strange particles all have lifetimes of the order of 10^{-10} sec or longer; i.e., more than 10^{+13} cy of nuclear time.

Let us write down the various invariance properties of these interactions besides invariance with respect to space time transformations. For all interactions it is found that the number of nucleons is conserved, and that charge is conserved. If we however limit ourselves to the strong and the electromagnetic interactions, we find that there is another conservation law, which one refers to as the conservation of "strangeness." If one makes the assignments of strangeness as follows:

$$n,\ p,\ \pi^+,\ \pi^0,\ \pi^- \qquad S=0$$
$$K^-,\ \bar{K}^0,\ \Lambda^0,\ \Sigma^+,\ \Sigma^0,\ \Sigma^- \qquad S=-1$$
$$\Xi^- \qquad S=-2$$
$$K^+,\ K^0 \qquad S=+1$$

then the experimental production data all conform to the statement that the strong interactions conserve strangeness. Thus, for example, in

$$\pi^-+p\rightarrow K^0+\Lambda^0,$$
$$n+n\rightarrow\Lambda^0+\Lambda^0,$$

the first reaction is expected to be a strong interaction, and the second one should be forbidden as a strong interaction. These indeed have been borne out experimentally. Finally, if we confine ourselves to the strong interactions only, the isotopic spin seems to be a good quantum number, just as in pion physics and in nuclear physics for light nuclei. One of the components, I_3, is related to the charge Q, the heavy particle number N, and the strangeness quantum number S in the following way:

$$I_3=Q-\frac{N}{2}-\frac{S}{2}.$$

The concept of strangeness has proved to be extremely fruitful in the past year or two, and although we do not understand the origin of this conservation law, nor do we know what new additional selection rules and quantum numbers will turn up in the future, it seems fair to say that strangeness conservation is here to stay.

As to theoretical problems of current interest: first there is the question of $\theta^0-\bar{\theta}^0$, (or $K^0-\bar{K}^0$). It was pointed out a few years ago that since the θ^0 particle has a strangeness of $+1$, its antiparticle $\bar{\theta}^0$ must have strangeness -1 and must therefore be a different particle. The decay of these particles was then expected to exhibit rather queer properties. This is of great interest, because these properties of the decay were predicted purely on the basis of symmetry principles. An experimental test of these predictions is in progress. Another problem which has been occupying a number of physicists is the question of the possible methods for the determination of the spins of the various strange particles Then there is the question: just how do the weak interactions violate strangeness? Also there is the study of hyperfragments, which are nuclei with Λ^0 particles bound in them It has been pointed out that an examination of the binding energies of light hyperfragments will yield some information about the Λ^0-nucleon force such as its spin dependence Lastly there

TABLE II. Experimental knowledge about neutral and positive mesons.

Meson	Decay product	Mass	Mean life	Fraction of all K^+
θ^0	$\pi^++\pi^-$	See below	1.7×10^{-10} sec	
$\theta_2{}^0$	$?+?+?$			
$K_{\pi3}{}^+\equiv\tau^+$	$2\pi^++\pi^-$	All K^+ and K^0 have mass $\sim966m_e$	All K^+ about $1\ \times10^{-8}$ sec	6%
$\tau^{+\prime}$	$2\pi^0+\pi^+$			2%
$K_{\pi2}{}^+$	$\pi^++\pi^0$			25%
$K_{\mu2}{}^+$	$\mu^++\nu$			60%
$K_{\mu3}{}^+$	$\mu^++\nu+\pi^0$			~3%
$K_{e3}{}^+$	$e^++?+?$			~5%

[3] See, e. g., A. M. Shapiro, Revs. Modern Phys. 28, 164 (1956).

is the subject of the so-called "freak" particles, which have turned up periodically, and which do not seem to fit into any established categories These and other questions have been discussed in the literature

I now turn to the last topic, namely, the question of the identity of the K particles Interest in this subject started when it was pointed out that by studying the decay of the $K_{\pi 3}{}^{+}$, also known as the τ^{+} meson, into three charged π mesons, it is possible to conclude something about its spin and parity Let me give you a simple example of the kind of arguments involved. Suppose one of the π mesons resulting from the decay of a τ meson is very slow, so that one may assume that it comes out in an s state relative to the center of mass of the other two π mesons. What can one conclude under these circumstances? Well the remaining two pions will be in one of the states $0^{+}, 1^{-}, 2^{+}, \cdots$, because the pions are spinless, and their intrinsic parities add up to 'even', so that the spin and parity of the two pion system are identical with those of the orbital part of their wave function. The extra π meson is odd, and contributes no angular momentum; we can therefore conclude that the spin and parity of the 3 pions together can only be 0^{-} or 1^{+} or 2^{-} $\cdots$. Since spin and parity are conserved in any process, the τ meson spin parity also must be $0^{-}, 1^{+}, 2^{-}$, etc. The $K_{\pi 2}$ by the same argument can only be $0^{+}, 1^{-}, 2^{+}, \cdots$. Therefore we may conclude that if ever the τ meson is found to decay into three pions one of which is slow, the $K_{\pi 2}$ and τ cannot be the same particle. Of course, the analysis is not quite so simple, because there is always the question: how slow is "slow" for the decay pion? To investigate this problem, the distribution of the momenta of the decay products of the τ has been plotted in a triangular plot and the full distribution studied against the various possible spin parity assignments for τ. With the one-thousand or so points available at the present moment the general opinion is that the τ is not the same particle as the $K_{\pi 2}{}^{+} \equiv \theta^{+}$.

However it will not do to jump to hasty conclusions. This is because experimentally the K mesons seem all to have the same masses and the same lifetimes. The masses are known to an accuracy of say from 2 to 10 electron masses, or a fraction of a percent, and the lifetimes are known to an accuracy of say 20%. Since particles which have different spin and parity values, and which have strong interactions with the nucleons and pions, are not expected to have identical masses and lifetimes, one is forced to keep the question open whether the inference mentioned above that the τ^{+} and θ^{+} are not the same particle is conclusive. Parenthetically, I might add that the inference would certainly have been regarded as conclusive, and in fact more well founded than many inferences in physics, had it not been for the anomaly of mass and lifetime degeneracies.

There has been in the past year a number of speculations concerning the reason for this anomaly. There was the suggestion that the equality of masses of two particles, say τ and θ, which have the same spin but opposite parity might come about in the following way: consider the process $\theta^{+} \to p + \bar{\Lambda}^{0} \to \theta^{+}$: the θ^{+} particle virtually dissociates into a proton and an anti-Λ^{0}, which then recombine. Such a process contributes to the mass of the θ^{+} particle, and does so appreciably, because all the interactions involved are strong interactions. Consider the same process for the τ^{+} particle: $\tau^{+} \to p + \bar{\Lambda}^{0} \to \tau^{+}$. Because of its different parity the contribution to the mass of τ^{+} from this virtual dissociation will be different (as the virtual particles will be in different orbital states relative to one another). After looking at this for a while, it appears that the only way for the contribution to be equal is that there exist two Λ^{0} particles: $\Lambda_{1}{}^{0}$ and $\Lambda_{2}{}^{0}$, which have opposite parities, so that the process $\theta^{+} \to p + \bar{\Lambda}_{1}{}^{0}$ has its counterpart in $\tau^{+} \to p + \bar{\Lambda}_{2}{}^{0}$ with the same orbital motion. The proposal has therefore been made that there may exist another symmetry operation called "parity conjugation," C_{p}, which leaves the strong interactions invariant, and which has the property of interchanging θ and τ. Also it is assumed that there are two kinds of Λ^{0}'s, and two kinds of Σ's, and that they go into one another under the operation of C_{p}.

It can easily be shown that the weak interactions which violate strangeness conservation cannot be invariant under C_{p}. In fact since $\Lambda_{1}{}^{0}$ and $\Lambda_{2}{}^{0}$ do not have the same parity, they decay in quite different ways into their final products. For example, the final product of the decay

$$\Lambda^{0} \to \pi^{-} + p$$

would in one case be in even relative orbital states, and in the other case be in odd relative orbital states, and therefore there is no reason why the two Λ^{0}'s should have the same lifetimes. This is also true for the θ and the τ. Thus the parity conjugation concept does not explain the equality of lifetimes.

An obvious test whether this speculation is correct is to see whether there do indeed exist such "parity doublets." One could for example look for evidences for two lifetimes for the Λ^{0}'s or the Σ's. Another way of detecting parity doublets is through the interference of the two decay modes of the two members of a parity doublet This will work if the mass difference between the two members of the doublet is very small, say of the order of 10^{-5} ev. In that case, let us consider an experiment currently being done in Berkeley: the capture of K^{-} particles in a hydrogen bubble chamber, and single out the process

$$K^{-} + p \to \Sigma^{-} + \pi^{+}.$$

What is seen in the experiment is a K^{-} coming to rest in the chamber, and a π^{+} and a Σ^{-} coming out. The Σ^{-} is relatively short-lived, and most of them decay before coming to rest, as follows

$$\Sigma^{-} \to \pi^{-} + n.$$

Let us now analyze the angular distribution of the decay particles (say the pion) under the assumption that the spin of Σ is $\frac{1}{2}$. The angle in question is the angle θ between the decay π^- and the line of flight of Σ^- in the rest system of Σ^-. We can analyze the spin into "up" and "down" components along the direction of motion of the Σ particle, and after doing this, let us make a reflection in the plane containing the lines of flight of the Σ^- and the π^- coming from its decay. Since under this reflection the spin flips (to see this picture the spin as a rotating top, so that under the reflection the sense of rotation changes, and so does the spin), and it is assumed that under a reflection the interactions are invariant, the Σ^- must have equal populations of up and down spins. In other words it is unpolarized. The angular distribution with respect to θ must then be spherically symmetric for ordinary Σ^-. However if the Σ^- does not have a definite parity, i.e. if it consists of a Σ_1^- and a Σ_2^- having opposite parities, then under an inversion they still both flip the spin. But due to their different parities they acquire a 180° phase difference between them, and therefore they would interfere, giving a forward-backward asymmetry (i.e., an asymmetry between θ and $\pi-\theta$,) a conclusion that can be experimentally tested.

With the existence of parity doublets of nearly equal masses one expects some very interesting phenomena. A beam of spin $\frac{1}{2}$ particles is usually described by four real parameters: its intensity, and a polarization vector which is the average spin of the beam. For a beam of spin $\frac{1}{2}$ particles with parity doublet structure the description is vastly more complex, because you now need four real intensities, one for the intensity of the particles of one parity, one for that of the other parity, and two for the real and the imaginary parts of the interference intensity. Also, whereas only one space vector is needed to describe the polarization in the usual case, now four space vectors are needed to describe the polarization: namely one describing the average spin of one parity one for that of the other, and two describing the interference polarization. The interference intensities are pseudoscalars instead of scalars, and the interference polarization vectors are vectors instead of axial vectors. The recognition of these transformation properties leads immediately to a number of conclusions about interference effects of parity doublets. These can all be subjected to experimental tests, and we may hope to have a clarification of the situation in the next few months.

A much more drastic thought has been advanced regarding this problem, namely the observation that both the mass identity and the lifetime identity can be understood if parity is not conserved in the decay. There need then be only one K particle, which has a definite parity upon production, but which can decay into various parities. This again is a proposal which can be subjected to experimental tests. The first question which comes to mind is: is the present evidence for parity conservation good enough to rule out such a proposal? One finds that the most accurate experiment to date, which tests and confirms the conservation of parity in strong and electromagnetic interactions, is the measurement of the electric dipole moment of the neutron which was found to be accurately zero. This experiment, though extremely accurate, is however not sufficient to yield a conclusive statement about whether the weak interactions also conserve parity, because the weak interactions are just too weak to influence the experiment. Furthermore, an examination of the tremendous bulk of knowledge about the beta interaction shows that to date there is no evidence there either for or against parity conservation in the weak interactions.

Fortunately, one can propose experiments which can test parity conservation in weak interactions, experiments which essentially test whether the weak interactions show a preference of one-sensed screw to the other. To give a typical example, consider the process

$$\pi^- + p \rightarrow \Lambda^0 + K^0$$

and examine the π^- coming from the decay of the Λ^0,

$$\Lambda^0 \rightarrow p + \pi^-.$$

Let us now ask whether the decay product π^- has an average component of momentum perpendicular to the plane of the incident π^- and the Λ^0, this plane being defined in a particular way, say, by having the Λ^0 direction always to the right of the incoming π^-. If there is such an average component, it would define a sense of screw, and its existence would indicate a non-conservation of parity in one of the processes involved, which would have to be the decay process, since the strong interaction is invariant under inversion. Another way of putting it is this: if the momentum of the incident π^- is $\mathbf{k}_1$, that of the Λ^0 is $\mathbf{k}_2$, and that of the outgoing π^- is $\mathbf{k}_3$, what we are looking for is the distribution with respect to the sign of $(\mathbf{k}_1 \times \mathbf{k}_2) \cdot \mathbf{k}_3$. This expression changes sign under space inversion. Unequal distribution with respect to its sign therefore means violation of invariance under inversion.

Another experiment, which is perhaps more easily done, is to study the angular distribution of the beta decay of an oriented nucleus. What is involved is to see whether there is an asymmetry with respect to the sign of $\boldsymbol{\sigma} \cdot \mathbf{K}$, where $\mathbf{K}$ is the electron momentum, and $\boldsymbol{\sigma}$ the spin of the nucleus. I understand that this experiment is currently being done at Columbia, and at the National Bureau of Standards.

When one opens the question of whether parity is or is not conserved for the weak interactions, one must also raise the question of the validity of the other conservation laws. Let us list the various conservation laws: they are the conservation of energy, momentum, angular momentum, parity, the invariance under time reversal (T) and the invariance under charge conju-

gation (C). Energy and momentum seem to be conserved in the weak interactions. The analysis of the conservation of angular momentum in weak interactions is more complex than the corresponding analysis of parity conservation, because parity conservation can be violated in only one way, whereas angular momentum conservation can be violated in a large number of ways. To my knowledge this analysis has not been carried out. Let us now consider T and C. A closer look at these conservation laws shows that there is no proof that these are conserved in weak interactions. One might at first think that the equality of the lifetimes of the π^+ and π^-, or the μ^+ and μ^-, which are charge conjugate particles, says something about this matter. But an examination of this problem shows that the lifetimes will be equal to the lowest order in the weak coupling constant (which is to an order of accuracy of 10^{-12}, and therefore essentially infinite accuracy) merely on the assumption of invariance under orthochronous Lorentz transformations (i.e., Lorentz transformations in which there are neither space nor time inversions). Because of this, the absolute invariance under C must be regarded as experimentally not proved.

Let me conclude with the point that if parity is indeed not strictly conserved, there would be a preference for either right handedness or left handedness in the universe, and this would appear to be unaesthetic. One may however observe that if one's definition of invariance is generalized, the question may appear in a quite different light. Let me give a simple example which T. D. Lee and I have speculated about: suppose that parity is not conserved, i.e., that there is a difference on going from a left-handed to a right-handed coordinate system. It may turn out, however, that by simultaneously going from one coordinate system to the other *and* switching to the anti-world, i.e., replacing π^+ by π^-, protons by antiprotons, etc., symmetry is regained. This merely shows that there is a richness in the structure of these symmetry laws which we are quite far from comprehending.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 2 APRIL, 1957

On the Origins of Cosmic Rays

PHILIP MORRISON

Department of Physics and Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York

IN 1954, Professor Einstein once remarked that there were two easily observable phenomena that, in his opinion, showed a deep fundamental lack in our knowledge of the physical world. These, he said, were the cosmic rays, and the terrestrial magnetic field. We can be all but sure today that he was wrong in that intuition, but he was expressing exactly that feeling which theoretical physicists have had for many decades, which somehow justifies the inclusion of topics like this one in a program on theoretical physics.

It is my opinion that they no longer quite belong in such a discussion. They have to do neither with the fundamental laws of physics nor with the difficult problems of applying the fundamental laws to relatively simple systems. On the contrary, both of these phenomena (which turn out to have the same fundamental origin) are examples of application of good old classical physics to a world of atomic materials, using nothing but the still unknown solutions to the nonlinear equations of magnetohydrodynamics.

The beginning of the path toward an explanation was shown to us perhaps 20 years ago when Fermi, Vallarta, and others began their series of investigations into the effect of the earth's field upon the cosmic rays which has led us now into a nearly complete understanding of, say, the latitude effect. The idea was gained, and given quantitative support, that the cosmic rays were not given and immutable, something whose origin could not be approached because they were so far beyond ordinary experience. This idea was aided by that kind of theoretical result, but most especially by the facility with which experimental physicists have *made* cosmic rays. For they have the same problems and the same difficulties that are encountered in nature in the large.

The study of the origin of cosmic rays (while we do not yet have a fully understandable, and certainly not a quantitative model) is no longer devoted mainly to the interpretation of empirical evidence obtained from properties of the cosmic rays themselves, but must draw as well from general astrophysical means of observation. These wider results are for the first time sharply relevant this year, and therefore it is perhaps appropriate to summarize very briefly what sort of path one can take for a tentative understanding of this extraordinary and interesting phenomenon. I intend to describe just one path; I know that there are many; I know that there are even people in the audience who have produced other paths for the explanation. The one I shall point out is the most coherent we have today. It is still a model;

it depends upon fitting together loosely many ill-defined parts, and naturally the change of any one of those parts can seriously affect the coherence of the model. Nevertheless, I think it is the best we have at present. As far as I can see, it has no clear inconsistency with anything we know, and it unites a great many phenomena which otherwise would have to stand as separated.

The path which I shall outline is associated with the names of Alfvén, Fermi, Cocconi, the Moscow group (Shklovski and Ginzburg), and Hoyle and Burbidge. I shall try to describe very briefly the logical structure of the ideas, the elements of the theory, but only some of the numbers and the specific mechanisms which bear on it.

The principal facts with which one begins are like many facts we have heard about in recent sessions. They are not at all facts, but only approximately true and simplifying statements. The two first principles are the isotropy and the time independence of the cosmic radiation. It is grossly isotropic, and grossly time independent. That both the isotropy and the time independence fail in detail is clear. Indeed their very failure has given us perhaps the clearest clues to the mechanism which may account for the over-all isotropy and constancy.

The most spectacular failure of both of these principles occurred on February 22, 1956, when the cosmic-ray intensity shot up by a factor of 2 or 3 even for hard mesons at sea level and by a factor of 100 for the neutron component. An analysis of this event and of a few others like it has given a pretty clear understanding that the cosmic rays do not come to us through empty space. They may originate, as these extra particles do in this particular case, on the sun, where the event can be seen to be in sharp time coincidence with extraordinary events on the surface of the sun. But whether or not they originate there, they pass to us through space which is by no means empty. It is the discovery of the population of this "vacuum" with physical, not simply virtual particles, which was the real progress of the past few years, and which has made possible an understanding of the cosmic rays.

The precession radius of curvature is given in familiar units by $R=E/300B$, the radius R in centimeters, the energy E in ev and B in gauss.

Observations have shown directly that the geomagnetic field affecting cosmic rays incident on the earth is skew from the dipole field evaluated on the surface of the earth, by about a tenth of a radian. It seems hard to doubt that this can only be due to the summing of the earth's field with some external fields out in space, not produced by the internal currents which make the dipole field of the earth. Since the deflection of a tenth of a radian comes in at about ten earth radii, what sort of field might be effective still in bending cosmic rays, that is, in producing cosmic-ray deflections with R above ten earth radii for a typical cosmic ray of 10 Bev energy? This turns out to be about one milligauss or a little less.

From this it is pretty clear that in the region around the earth, say ten radii away, there is a field (which may be transient, but which is there sometimes) of the order of a milligauss, perhaps. In this field the radius of curvature of a cosmic ray amounts to one light second. The light transit time from the sun, from which the special pulse of cosmic rays originated on February 22, 1956, is eight minutes, so you see that such a field would mean that the path from sun to earth could by no means be a straight line. Indeed, independent analyses of such flare events have given fair confirmation to the idea that, at least under these peculiar circumstances, the earth-sun region is by no means empty for the cosmic rays, but rather the cosmic-ray path must be some more or less chaotic random walk in this region, the diffusion being caused by scattering, not from nuclear collisions, but from collisions with more or less coherent magnetic fields extensive enough and strong enough (from 10^{-5} up to 10^{-3} gauss) to induce radii of curvature of magnitude small compared to the earth-sun distance. Then a diffusion-theoretic treatment, using elementary ideas of diffusion theories and not implausible ideas of the geometry of such transient clouds of scattering magnetic centers that might come from the sun, gives a semi-quantitative fit to the decay time of a solar flare, to the spectral changes, and to a number of other features.

I take this, not as a demonstration of the origin of the cosmic rays at all, but as an example of the fact that there can exist, in more or less empty space, (to be sure, this is space very near a star of the main sequence, and not typical galactic space) a kind of magnetic material that prevents cosmic rays from traveling in straight lines. It is exactly this which is the key to the whole phenomenon. Astronomers have independently obtained other evidence for fields in galactic space.

The gross cosmic-ray isotropy and time constancy define our first major point; the second major point, the flare events and the skew field, argues for the modification of cosmic-ray direction when they propagate in "empty" space. The third main property derived from the rays themselves is their distribution in energy and charge. There is a paucity of cosmic rays below a couple of Bev, a low-energy "cutoff" which varies with the solar activity cycle of 11 years, and is one more sign of the influence of the magnetic regime around the sun upon our cosmic-ray flux. From some 10 Bev all the way to 10^4 Bev, and with less sureness but still quite plausibly up to 10^7 or even 10^8 Bev, the cosmic-ray integral energy spectrum follows a simple power law at least in rough approximation, with

$$N(>E)=\text{const}/E^{\alpha}, \quad \text{with} \quad 1.3<\alpha<1.8, \qquad (1)$$

where $N(>E)$ is the flux of primary rays with total energies greater than E. The exponent changes slowly over the energy range, not going beyond the limits shown over pretty much the whole range. Besides this

over-all energy spectrum, we have some information, though much less, about the charge distribution among the primaries. Not only protons enter the atmosphere, but helium, elements of the CNO group, and some nuclei as heavy as that of iron. Protons predominate, the others being present in amounts very roughly like those anticipated for the mean gas sample of the universe, with maybe a tenth of all the flux being in alpha particles. The energy spectrum of these heavy primaries is poorly known, but for energies up to some 20 Bev per nucleon, it is like that of the protons, at least roughly, and there is some sign that this similarity extends to much higher energies. One must be prepared for departures from the power law of energy distribution, want of close parallelism between the heavies and the proton spectrum, and all sorts of other complications. But the general picture here appears to apply up to about 10^8 Bev or beyond.

About the only additional fact one needs to know for a discussion on this level of accuracy is the total energy density for various kinds of energy in space. The mean energy densities are for various forms, in a typical galactic region not near a star, in units of ev/cc as shown in Table I. This sets the level of the energy to be

TABLE I. Energy densities in space.

Type	ev/cm³
Cosmic rays	1
Starlight	1
Turbulent gas motion	1–10
Kinetic energy of rotation for galaxy as a whole	1000

accounted for, and provides a few comparisons. A typical observer somewhere in a galactic spiral arm would presumably see such values whose variability is indicated by the roundness of the figures.

We see the striking fact that the cosmic ray energy density is like that of starlight. Yet starlight pours out of every star, thermodynamically guaranteed as the degradation product of thermonuclear reactions, and cosmic rays do not as far as we know pour out of any ordinary stars at all. It is true that some do come from the sun, but the rate at which they are made, if extrapolated to all the stars of the system, would amount to a tiny fraction of this energy density (10^{-8} perhaps). Even if there are some unusual stars which flare up all the time, they cannot make up this enormous discrepancy. Moreover, the sun makes very few cosmic ray particles above 10 Bev in its most violent spasms, and stars not very different from the sun are not likely to make some rays 10^7 times as energetic. The turbulent energy looked for a while like a hopeful source, but if it did supply the cosmic rays, they would be the heaviest drag on the turbulence, and the hydrodynamics of the galactic gas would be determined by the cosmic-ray flux. This now seems unlikely, for various reasons. The rotation energy looks like a nice hope. There's plenty of energy there, but it is not easy to convert the kinetic energy of stellar systems whirling about the galaxy in circles of 10 000-light-year radius into kinetic energy of single protons moving in a chaotic motion; everyone has stopped trying to derive any comfort from that big number. Therefore, I will maintain now, and in the sequel will try to demonstrate, that the origin of much of the cosmic-ray energy is, like that of starlight, nuclear in character. This is agreeable, because nuclear energy provides a source of energy that we know how to calculate, and maybe it is true.

What are the processes? This is a kind of analysis or summary of the processes which people have used to describe the mechanisms. I am going to list a few rather formal points and then try to describe them in turn, to show what kind of model can be built up: stirring, storage and loss, acceleration, injection, and cutoff.

Stirring is that process which makes the cosmic rays isotropic, and which arranges them to be time-constant as far as possible. These must be extrasolar processes; it is very hard to doubt that the stirring is caused by a chaotic magnetic field, extending through some or all of the galaxy. Just where it resides I will try to make a little clearer by some photographs, but its presence, I think, we will take as given; nearly all theories join there. The only way in which theories differ is in how these stirring fields are constructed: tight enough to prevent leakage out of the ends, or chaotic; flat or round or what they may be; but everybody has a stirring field. Otherwise, one ought to see where the sources are, for it is not plausible that the sources are isotropically distributed about the earth. This is the Copernican assumption of the cosmic-ray theory.

Next consider storage and loss: this is the root idea in a way. It goes with the stirring. If there is stirring, then the path length of a cosmic ray particle is much longer than a straight line from its source. That implies the energy density may be high without the rate of energy production being very high, because there is a pool of cosmic rays formed by storage. No such pool is formed for light photons; they must move in straight lines, since the galaxy is thin to them. The apparent equality of cosmic-ray and starlight energy then does not mean that one must look for a cosmic-ray source as intense as the light source. This cosmic-ray source, it turns out, needs to be only about 10^{-4} or 10^{-5} as intense, which is already hard, but at least is not so impossible as if it were one to one. The numbers 10^4 to 10^5 come from a comparison of the successfulness of the stirring, that is, the degree of isotropy, with the loss processes which must limit the storage eventually. The storage must be limited. The cosmic rays cannot be arbitrarily old in the galaxy: (a), because they would make nuclear collisions; and (b), because if they escape nuclear collisions, it is extremely hard to define a trapping magnetic field that will hold particles in indefinitely (as all designers of particle accelerators will

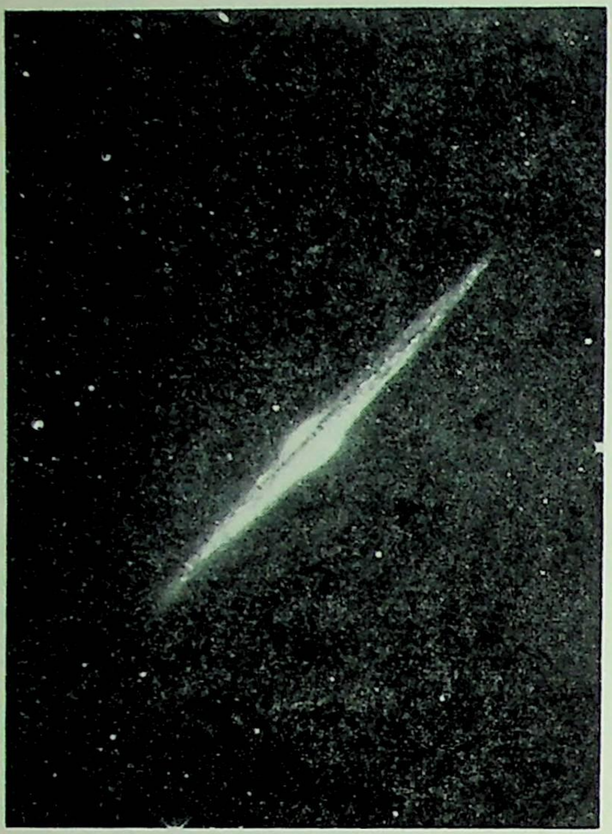


FIG. 1. A spiral nebula seen edge on. A galaxy of the Coma Berenices cluster, photographed at Mt. Wilson, NGC 4565.

agree), and certainly with turbulent and chaotic fields it must be even harder. So the rays are going to leak out —that means moving from a region where the energy density of cosmic rays is the tabulated value to an (assumed) region where it is smaller. What that region is we discuss later.

On the other hand, the rays might be destroyed by nuclear collision. One knows that they are not destroyed overwhelmingly by collision because we do see heavy primaries, which means that even if the particles started out all iron (not at all impossible, but not plausible) still they could have reasonably traversed not more than a few tens of grams per square centimeter, a few mean free paths. And they may have moved through much less than one gram per square centimeter. (The argument to be settled by the presence or absence of Li, Be, B, elements which are missing in the chemical abundance of the universe as a whole, is an argument which bears on this question less sharply than it once did. It bears rather on a distinction between less than half a gram and more than a couple of grams.) I think everyone will agree that the cosmic-ray beam does not come through more than five or ten grams/cm^2 anyhow.

The mass of a one cm^2 column of matter along the galactic diameter is a couple of grams. The cosmic rays have not gone through such gas for a distance more than 10 or 20 times that, certainly not any enormously greater distance. Of course, such storage paths guarantee long life, and guarantee that rapid changes of source strength are not likely to be reflected in a change of the intensity observed where we sit, because the pool has stored up rays over a long time. Only if we are so lucky as to live near an actively-generating source of cosmic rays will we see a little fresh injection, as we do. That little contribution is from the sun.

Where is the storage? This is the first and the hardest question to answer. The erroneous answer which we once gave is notorious. I am now giving you another answer. I hope it is not as bad.

Figure 1 shows what led us into giving the wrong answer. This is a photo of a typical spiral galaxy. It is not our own, but it is a spiral of similar class to ours. This thing is characteristically a disk. It is very easy to say, "The cosmic rays are stored in that flat disk, and therefore they are going to be lost wholly by leaking out of the flat faces." This gives rise to a serious problem, because then the storage time is small, and you are forced to accelerate the cosmic rays very quickly, since this disk is so extraordinarily thin.

Figure 2 is M31 in Andromeda which is a sister galaxy, just like ours, seen a little edgewise. Here there is a tilted-disk effect. It is also quite a flat structure. This is a typical picture on a Palomar photographic plate.

In Fig. 3 various distributions through the Andromeda nebula are plotted against angular distance from the center, looking only more or less perpendicular, not correcting for our oblique viewpoint. The lower curve is a curve of the mass distribution, as calculated from the distribution of stars. It is quite a thin disk. The higher broad curve on the other hand shows no sign of a core plus a cloud; it is now just a big near-sphere. This is the picture that radioastronomers see at a frequency of a couple of hundred megacycles. So there is something,



FIG. 2. Our neighbor galaxy, the spiral NGC 224, the Great Andromeda Nebula. In the optical telescope it appears as a tilted flat disk of stars.

invisible to the eye but a good source of radio radiation, making a spiral galaxy nearly spherical.

The proposal that we do away with disk storage and store the cosmic rays in the sphere is extremely likely to be true; but I would like to defer the argument as to why the presence of a radio signal from the whole sphere of Andromeda is a strong argument in favor of storage of cosmic rays in that spherical volume. To have thus increased the storage volume from that of a disk to that of a sphere is very agreeable. It gives a hundred times larger storage volume which means that, since the radio source density does not change much with radius it is fair to take roughly a hundred times the total cosmic-ray energy. Fortunately the lifetime goes up also by a hundred times or two, because there is so much more space in which the rays can make the random path. This means the rate of production is about the same, perhaps a little bit less. But the time rate of acceleration, the rate of storing energy in the particles, does not need to be sizeable. There is plenty of time. Particles can come back and forth many times; or can be accelerated in quite a different way. The difficulty of doing all this in a flat and leaky disk, is gone. Generally speaking, everyone is happy with the astronomical evidence that the cosmic-ray galaxy is not (as it looked to the eye) flat, but as it looks to the radio antenna, spherical. The gas is so tenuous that nuclear collision remains a minor cause of loss.

Next, consider acceleration. This is the process about which we know the least. It was originally proposed, I think by Fermi, that this very chaotic scattering would add energy to the cosmic-ray particle as it changed its direction in collisions. This is a process strictly analogous to the moderation of neutrons as they scatter elastically in a low-temperature moderator. In this case, however, the scattered particles gain energy at the expense of the energy of the large magnetic clouds which cause the scattering. One may say that these large clouds moving at velocities of some tens of kilometers per second and possessing masses of stellar size have

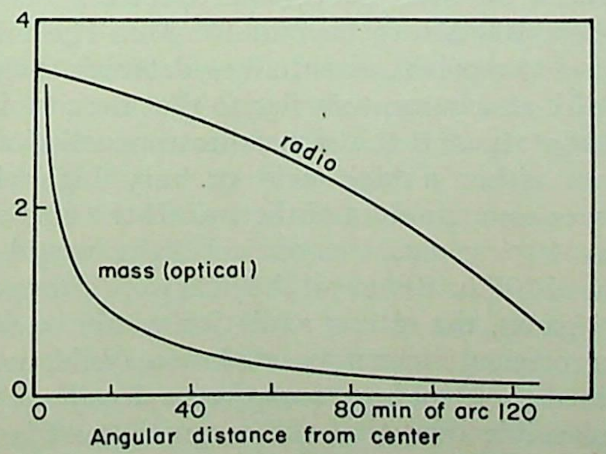


FIG. 3. Distributions seen in the Andromeda galaxy, NGC 224. The upper curve shows the radio source intensity at 200 Mc, plotted on an arbitrary scale, as a function of angular distance from the center of the galaxy. The lower curve shows the mass density calculated from the optical observations on stars, in a similar graph [after J. E. Baldwin, Nature 174, 320 (1954)].

enormous "temperature:" i.e., large kinetic energies per translational degree of freedom. The cosmic-ray particles try to moderate against such clouds, but in moderating, they must of course rise in energy rather than decline as the neutrons do against cold moderator. It is clear from the magnitudes involved that equilibrium can never be reached, and that the highest energies attained by cosmic-ray particles will be limited by the time available for the process of inverse moderation. Fermi originally suggested that nuclear collisions would limit the process. This would strongly discriminate against heavy primaries in the beam, and seems to be excluded. This is perhaps the major consequence of the measurements on heavy primaries. The similarity in energy spectrum between hydrogen and the heavy primaries seems to imply that the process is limited, not by nuclear collisions which vary strongly as the mass of the nucleus varies, but by leakage out of the magnetic region responsible. This would be nearly independent of the type of particle involved.

The beauty of the Fermi mechanism in general is that it leads to a power-law energy spectrum in the simplest and most natural way. We may formulate the Fermi theory in the most naive form by writing the simple relations

$$dE/dn=\alpha E;\quad E=E_0e^{\alpha n};\quad \text{and}\quad dN/dn=-N/\bar{n}, \tag{2}$$

where α measures the relative energy gain in each collision with a scattering center, n is the number of collisions a given particle has made since its injection into the process, E is its total energy, and $\bar{n}$ the mean number of collisions before loss occurs. $N(n)$ is the flux of particles having made n collisions. With the simplifying assumption that both α and $\bar{n}$ are constants, independent of n, we obtain the spectrum by eliminating n:

$$N(>E)=\int_{n(E)}^{\infty} N(n)dn=\text{const}/E^{1/\alpha\bar{n}}, \tag{3}$$

where $N(>E)$ is the integral flux as in (1). Several detailed models have been proposed for particle motion in a magnetic region, all of which retain the feature that the energy goes up exponentially with collision number. This feature is pretty well maintained even when the mean energy gain is negligible, and only fluctuations can be counted upon. The combination of such an energy gain with the exponential number loss rate characteristic of a particle beam in almost any diffusion process gives a natural power law.

In (3), the exponent $1/\alpha\bar{n}$, which empirically is close to unity, is the combination of two numbers, one depending on the motion of magnetic clouds, the other on their mean field strengths and spatial extent. It is not obvious that these should be related, and if they were really independent (as they would be if loss arose from nuclear collisions) it would be a miracle that the ratio should come out near unity. It will appear immediately that we cannot interpret the full extent of the energy



FIG. 4. The Crab nebula, Messier 1, NGC 1952. Photographed at Mt. Wilson in the light of $H\alpha$ (6300–6700 A filter)



FIG. 5. The Crab nebula viewed in a portion of the spectrum without emission lines, seen in the light of its continuum only. (7200–8400 A filter.)

spectrum as being due to a single simple Fermi mechanism, but rather as the result of the superposition of many such processes, each arising out of a different physical situation. This suggests that perhaps some deeper dynamical connection exists between the extent of the magnetic region and its local properties. It is for the future to uncover such a relation if it is present.

The low-energy spectrum with its tailing-off fits fairly well even with the elementary theory. But we are sure from the year-by-year change of the cutoff (which all but disappeared in 1954) that the residual magnetic fields within the solar system play a major role.

All this looks pretty good, and one would be led to say that the Fermi theory in a sphere, with the loss by leakage and not by collision, would more or less handle everything, if there had not been some very unpleasant observations, some hard facts, to cope with. The hard fact is this: synchrotron radiation by electrons in a magnetic field has many characteristic properties. One of the properties is, of course, its continuum—a continuum which has a peak at a characteristic frequency depending upon the strength of the field, and the energy of a radiating electron. If the electrons have a wide range of energies, there will be a wide continuous range of radiation given off from any source of such synchrotron radiation. This radiation is completely plane polarized, electric vector in a direction normal to the lines of force of the magnetic field, especially for relativistic electrons. The characteristic frequency ν_c is given approximately by

$$\nu_c \cong 1.4\gamma^2 B \text{ Mc/sec} \qquad (4)$$

for B in gauss; here γ is the relativistic energy-to-rest-mass ratio for electrons. It is also quite easy to calculate what the loss rate in energy is; it is rather a heavy space rate of loss.

Shklovski in Moscow about three years ago, and after him, a couple of Russian astronomers, and then Oort and now Baade have produced a whole series of predictions and, finally, observational confirmation which demonstrates that synchrotron radiation in enormous amounts is indeed emitted by one very conspicuous and very interesting object. That object is the Crab Nebula, Messier 1 or NGC 1952.

Figure 4 displays the Crab nebula as seen in the Hale telescope on Mt. Palomar, revealing a complex filamentary structure, shining mainly in the emission line H_α. This cloud measures about six light years across its long axis.

In Fig. 5 the same object is viewed through optical filters which pass no important emission line. There is still plenty of light, for remarkably enough what we see here is a hazy amorphous mass shining with a continuous spectrum, a continuum responsible for about nine-tenths of all the light from this object in the visible. It is a bluish continuum; almost nothing like it had ever been seen before. The spectroscopists who tried to analyze it in terms of free-free transitions in hydrogen reached absurd results (which everyone believed for a decade). They had to impute to the structure a mass much greater than anyone had ever seen for any star. Now, we know that this object is the radioactive and exploding debris of a supernova explosion in the year AD 1054. It was an unusual and unstable star, but its mass is not more than several sun's masses.

The Crab nebula is a very strong radio source. At 200 Mc it is the third brightest discrete source in the sky. Shklovski first made the bold assertion that the anomalous radio emission and the optical continuum were part of the same continuous spectrum, stretching from below 100 Mc to beyond 10^8 Mc. The intensity spectrum—two points, essentially—determines a power-law fit with an exponent similar to that for cosmic rays, somewhat steeper. If this is synchrotron radiation there is present either a high field or very high electron energies, or some product of the two. If the field were as small as 10^{-4} gauss, the optical light would imply electrons of 10^{15} or 10^{16} ev; if the field were stronger than say 10^{-2} gauss, the rate of radiation would be so great that the original supernova explosion could not have maintained the structure more than a decade or two. It seems plausible then that the magnetic field is near a milligauss. But the convincing evidence that this brilliant source of light, a thousand times more luminous than the sun, is really shining by synchrotron radiation, is given by the photographs of Fig. 6 (kindly supplied by Professor Baade).

Figure 6 shows the Crab nebula through the Hale telescope, exposed behind filters which exclude the emission lines of the filamentary structure. The arrows on the photographs show the direction of the electric vector passed by the Polaroid filter through which Professor Baade last year took this remarkable pair of pictures. Choose any feature which is prominent in one picture, and you will notice a difference in the intensity of the feature on the other picture which is quite marked. What we are seeing here is the local direction of the electric vector; that is the direction of the normal to the magnetic field producing synchrotron radiation. If you map out the successive directions of polarization of these wispy structures of continuum you are thereby mapping out the magnetic field in this object. No one, as far as I know, has suggested any conceivable way by which you can get 100% polarized light in an object of this kind anywhere except by synchrotron radiation, and I take this as the strongest demonstration: (a) that there are strong magnetic fields in some parts of the universe, and (b) that when there are strong magnetic fields, there may be relativistic electrons moving in them. Next, one can ask where these relativistic electrons come from, and what bearing this has on the cosmic rays.

Professor Oort has drawn the directions of the lines of force to see the shape of the magnetic field; it is a sort of turbulent field, but has no visible fine structure. That may be deceiving, but I don't think it likely, since at any point one looked one could see many directions of polarization if the field were chaotic on small scale, and this would wash the effect out. The presence of 100% polarization means that, at least in the outer portions of the nebula the lines of force are not very chaotic. That is, they are uniform on a scale of a tenth of a light year, or a few tenths of a light year, and a strength of the order of a milligauss. This is something prodigious from the point of view of magnetic fields. It is impossible to doubt that here too are many fast electrons; the total energy content that can be computed for them is quite high. If there are fast electrons, it seems now extremely hard to doubt that there are also fast protons, or heavy particles, made by the same processes. Since the electrons feel the drag of radiating this very synchrotron radiation, it is hard for them to gain energy, moving in paths of short radius. It is relatively easier to accelerate protons in such paths of short radius.

This certainly implies that the heavy-particle spectrum would be even less steep in energy than is the electron spectrum, and that would very well fit, roughly speaking, a mean energy of at least 10^{11} or 10^{12} ev for protons; it may very well be higher by 10^{2-3}. It is even plausible that the source of the fast electrons is simply the collisions of the heavy particles, protons and heavier, with matter in the interior of this object, so that the secondary electrons finally coming from decay of μ mesons are what actually radiate. On this basis these objects are the sources of very large amounts of cosmic rays. I think it cannot be doubted that here is one such source (but one may always doubt that this provides a major class of cosmic-ray sources).

We lose by this means, it seems, Fermi's natural explanation of the energy spectrum. If there are many such supernovae, one per century in our galaxy, the sum of these supernovae is what feeds the pool of cosmic rays, to fill the whole sphere of the galaxy. It seems rather hard to believe that their spectra would all be the same. There is no serious difficulty, if supernovae produce 10^{39} ergs/sec in cosmic rays. It seems quite possible that that would be enough to fill the whole sphere with cosmic rays to the appropriate density, at least between 10^{10} and 10^{14} ev.

What is the reason for the radio source in the sphere which surrounds the galaxy? This is again electron synchrotron radiation, perhaps from secondary electrons produced by cosmic-ray primaries, away out there far from the stars, but in weak magnetic fields where it is no longer possible for them to make optical radiation yet quite possible for them to make radio radiation. This establishes the consistency of the argument that the cosmic rays must be stored throughout the galactic volume, finally reaching the edge, and leaking out.

With all this I would like to remark upon the high-energy cutoff. Its absence is hard to understand. It

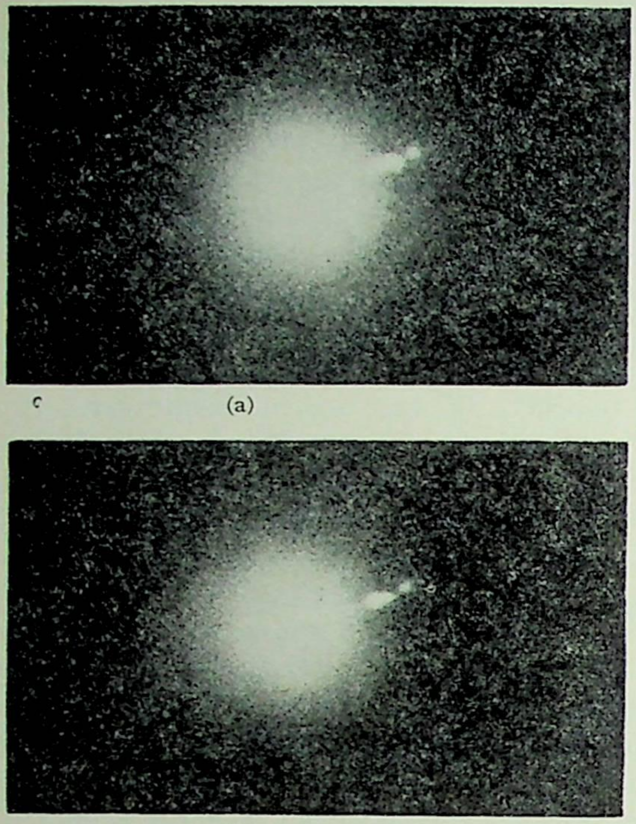


(a)



(b)

FIG. 6. The Crab nebula again seen by its continuous emission, but now through a polarizing filter in two perpendicular planes of polarization. The white arrows mark the direction of the electric vector passed by the filter in each case. Note how the structures tend to be aligned at right angles to the electric vector. The directions taken by the wispy structures at each point are the directions of the magnetic field there. (Photographs kindly supplied by Professor W. Baade of Palomar Observatory.)



(b)

Fig. 7. The bright central portion of the spherical galaxy M87 or NGC 4486, in the Virgo cluster. Note the broken linear structure extending up and right from the center. It is called the jet of M87. The two photographs are taken with differing directions of the electric vector, showing a marked polarization of the light from the different knobs within the jet [from W. Baade, Astrophys. J. **123**, 550 (1956)].

would be very nice if a cutoff were to be found. We thought, a couple of years ago, that we had good arguments why there would either be a failure of isotropy or a fall-off of intensity in the spectrum at 10^{15} or 10^{16} ev. Neither occurs.

Perhaps there is no cutoff in the spectrum because the source produces an abundance of fast particles; we can't be sure about that. Those high-energy particles that moved a long way through the galaxy and were kept in would still be of high energy, but those that leaked out into the general halo would have gone away from us; therefore even in this case anisotropy would be seen at high energy. The high-energy particles might come, say, from some spiral arm or even from the center of the galaxy, or from somewhere where there might be very special accelerating conditions. But there is no visible anisotropy, and there is no change in the energy spectrum so far up to 10^{18} ev. I say this with some misgivings; it is hard to interpret the giant-shower experiments which are the basis of the statement. Perhaps one is not fully certain of this; there are certain loopholes. But I think it is fair to say that one must be prepared to make cosmic rays about in the same intensity up to 10^{18} ev, maybe more. Now it is very hard to do that, because a particle of let us say 10^{17} ev, which is still to be accelerated by the factor 10 somewhere, would have a radius of curvature bigger than the Crab nebula even in the very strong fields of the Crab nebula. Certainly nothing in the Crab nebula is ever going to accelerate it by repeated events, if it cannot bend it even once. A linear accelerator with high electric fields over those distances is incredible.

How can one manage to avoid the cutoff? This is a rather dark question. A thousand-light-year mean free path in the galactic magnetic fields is what we have here; and a thousand light years takes us out of the galaxy in ten or a hundred steps. It is not very easy to accelerate in that time. There is, however, some hope. It is the usual hope of astrophysicists. When they learn of two things they don't understand, they say that probably these things are connected.

There is a large cluster of galaxies in the constellation Virgo, a cluster about twenty million light years away from us. A conspicuous member of the cluster, known as M87, is a sober-looking nearly spherical galaxy, as large as our own, but without a trace of spiral structure or disk flattening. It is a beautiful sphere of so-called Population II stars, very heavily populated in the center, and fading out to a thin haze of stars and star clusters a hundred-thousand light years from the center. In Fig. 7 the center portion of the object is shown in a rather short exposure. Note the extraordinary structure coming out of the central star mass; it is called the jet of M87. It is two- or three-thousand light years long, and a few hundred light years in diameter. Nothing like it is known in the heavens.

The two photographs (again from Professor Baade) show the object in two directions of polarized light. The knobs within the jet are plainly polarized. Measurements show about thirty or forty percent polarization along several directions. The polarized light is again a bluish continuum. It resembles nothing else ever seen except the light of the Crab nebula, but that light it closely resembles. The whole galaxy, not just the jet, is a strong radio source as well. These facts can hardly be interpreted otherwise than by saying that here too is a region of strong magnetic field, a great source of synchrotron radiation and of fast electrons which within the jet radiate the polarized optical continuum, and then leak out into the weaker fields which fill this whole galaxy, there to radiate in the radio region only.

The M87 jet is, however, built on a wholly different scale from the Crab nebula. Locally, it seems much the same, with perhaps a somewhat higher density of relativistic electrons. But it has perhaps ten million times the volume of the Crab nebula, and contains proportionately even more energy. Such a thing cannot arise in any part of the evolutionary history of a star. It may somehow obtain its energy from the collision of galaxies. We simply do not understand its origin. But there it is; its interpretation seems clear. Within it, assuming that it resembles the Crab except in scale, cosmic rays of energies up to 10^{21} ev or even higher might well be made, slowly to leak out into extragalactic space, and to us.

Perhaps such an object—one or a few might exist in most great galactic clusters—can explain the absence of

a high-energy cutoff. Our general picture of the cosmic-ray energy spectrum leads to the idea that increasing energy of the particle implies both longer time since its injection and greater distance to its birthplace. The bulk of the energy, at the low-energy end of the spectrum, reflects the long storage of these easily-contained rays in the local region of our galaxy, and their birth from stars and novae, perhaps, with some acceleration from the magnetic clouds of the galactic arms. The great range of energies from a few tens of Bev up to, say, 10^{14} ev or more, come from the pooled contributions of supernovae, with a density smoothed out over the whole spherical volume of our galaxy, not just the arms or the disk. Finally, with a spectrum very slowly decreasing with energy, but with a very low absolute density, the whole of extra-galactic space may be filled more or less uniformly with particles of energy up to well beyond the top experimental range. Isotropy is everywhere maintained by the diffusing fields; long storage times yield no major change in mass spectra, because the regions of storage contain smaller and smaller densities of matter as their volumes become bigger and bigger. The power law spectrum loses its simple origin, and decomes a smoothed-over statistical summation, with some still poorly-understood connection between local fields and extent of a magnetic region responsible for the nearly-constant value of the exponent.

One great question remains: how can the energy of a solar flare, a supernova remnant, or a galactic jet, be converted to cosmic rays with efficiencies of at least some percent (and indeed a thousand percent efficiency can follow from a naive faith in the data!). The answer to this problem, surely of hydromagnetic kind, is still only dimly seen. The solar flare is our nearest laboratory, and will receive close study in the years to come.

Now, I come to treacherous speculative ground, which I would prudently avoid had not Fred Hoyle gone there already. The working out of the evolution of elements makes clear that it is a mistake to ignore even the difficult paths pointed out to us by such a guide. Can any role be assigned in all this strange process to the possible presence of antimatter somewhere in the universe? There may be a sign or two that some such thing is involved. One very curious primary particle, of doubtful interpretation, has been seen. It just might be a nucleus of anti-silicon. The annihilation of anti-protons might play some role either in the injection of fast electrons or even just barely in the source of energy for the M87 jet. Surely it has no large place in the history of the Crab supernova. It is still uneconomical to postulate anti-matter, but it cannot be excluded either.

There are many holes in this discussion, but I will close as I began, with the remark that for the first time the origin of the cosmic rays must be discussed in a context which is so broadly related to a variety of astrophysical experience that we may think at last, even if the details are all wrong that the broad road is finally visible.

BIBLIOGRAPHY

Cosmic-ray theories:

H. Alfvén and N. Herlofson, Phys. Rev. **78**, 616 (1950). An early suggestion for the connection between radio stars and cosmic rays.

E. Fermi, Phys. Rev. **75**, 1169 (1949), and Astrophys. J. **119**, 1 (1954). Initial and improved versions of the basic theory.

G. Cocconi, Phys. Rev. **83**, 1193 (1951).

Morrison, Olbert, and Rossi, Phys. Rev. **94**, 440 (1954). Two discussions of the relation between diffusion, storage, and anisotropy.

V. L. Ginzburg, Nuovo cimento, Suppl. **3**, 38 (1956). A review in English of the theory of the Moscow workers, very much the basis of the present account.

Astronomical background:

For the Crab nebula:

W. Baade, Astrophys. J. **96**, 188 (1942). History of the star.

F. Hoyle *et al.* Phys. Rev. **103**, 1145 (1956). Nuclear processes.

J. H. Oort and T. Walraven [Bull. Astron. Inst. Netherlands **12**, 285 (1956)] analyzes the polarized light in detail.

I. S. Shklovski [Doklady Akad. Nauk SSSR **90**, 983 (1953)] predicts the polarization of the light.

For M87:

W. Baade, Astrophy. J. **123**, 550 (1956). Illustrations.

G. R. Burbidge [Astrophys. J. **124**, 416 (1956)] analyzes the polarized light and radio source.

I. S. Shklovski, Astron. J. Soviet Union **32**, 215 (1955) discusses and analyzes nature of object and light.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 2 APRIL, 1957

Stellar Energy Sources

E. E. SALPETER

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York

1. INTRODUCTION

A LARGE part of this discussion will be concerned with thermonuclear reactions and how they affect the evolution of individual stars. These reactions affect the life history of a star very greatly, since they produce most of the energy which feeds the star's radiation. Were it not for the nuclear energy sources, the vast majority of stars we now see would be black dwarfs. I will only give a brief outline of the present state of the art, which difficulties have been overcome and which difficulties remain. I will not attempt to name the many authors who have contributed to advances in the field.[1]

For energy production in stars, only a small number of thermonuclear reactions is of importance and we do not need to know their reaction rates with great accuracy. However, nuclear reactions in stars are also of interest in another connection: According to one of two rival schools of thought, all elements in the universe other than hydrogen were not made at the "beginning of the universe," but were made from hydrogen in stars by means of nuclear reactions. For testing this hypothesis one needs to consider a very large number of nuclear reactions in great detail and needs to know the ratios of various reaction rates fairly accurately. Since many papers in the last few years have concentrated on the question of "element cooking," I will at least mention these considerations in passing.

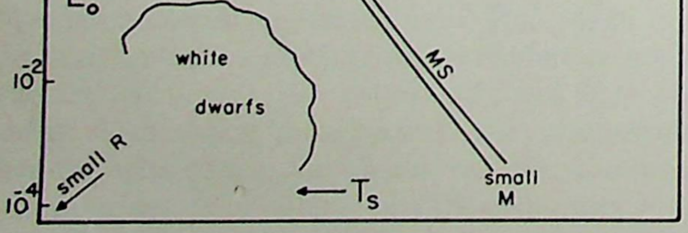


FIG. 1. Schematic Hertzsprung-Russell diagram for nearby stars of all ages. L is the luminosity, T_s is the surface temperature, M the mass and R the radius. *MS* stands for "main sequence." The circle denotes the position of the sun.

2. ASTROPHYSICAL OBSERVATIONS

The following is an oversimplified outline of some purely astrophysical observations relevant to stellar energy production and evolution: The most important are embodied in the Hertzsprung-Russell diagram, where one plots a star's absolute luminosity L (the total radiation emitted per second) against its effective surface temperature T_s, for a large number of stars.[2] Given L and T_s, the radius R of the star (or, more precisely, of its photosphere, the layer from which the observed radiation emanates) follows from the law of blackbody radiation. For a much smaller number of stars (too few, alas) the stellar mass M can also be obtained from observations.[3] The three quantities M, L, and R are the main link between stellar interior calculations and observation.

The main features of the Hertzsprung-Russell diagram—if stars of all ages, near and far, are included—are shown (highly schematic) in Fig. 1: First of all we have a fairly narrow band of points, the "main sequence" stretching from stars of low luminosity L and low surface temperature T_s to hot and luminous ones. To the right of the main sequence are the "red giants," stars of high luminosity and low surface temperature and hence with very large radii R. The red giants, unlike the main sequence, are scattered over a fairly large area of the Hertzsprung-Russell diagram. Finally, to the left of the main sequence, are the "white dwarfs," hot stars of low luminosity with very small radii.

We have fairly good statistics on the masses of main sequence stars (except the very luminous ones, which are rare): There is an almost one-to-one correspondence between the position of a star along the main sequence in an L–R diagram and its mass M, more luminous stars having larger masses (the luminosity increases very rapidly with increasing mass). White dwarfs are slightly less massive than the sun. Unfortunately we have very little reliable data on the masses of red giants, except that they are in general larger than the solar mass M_0. This paucity of quantitative data had led in prewar days to the "mass-luminosity relation," which is now thought to be partly incorrect: This relation supposed that any red giant had the same mass as a main sequence star

[1] The list of references is neither complete nor, necessarily, representative. The selection is biassed in favor of recent papers which include broader reviews of some topics or a large bibliography.

[2] (a) J. A. Hynek, *Astrophysics, A Symposium* (McGraw-Hill Book Company, Inc., New York, 1951), Chaps. 1 and 2 (b) Aller, *The Atmospheres of the Sun and Stars* (Ronald Press, New York, 1953); (c) R. D. Woolley and D. W. N. Stibbs, *The Outer Layers of A Star* (Clarendon Press, Oxford, 1953).

[3] See reference 2a, Chaps. 9 and 11; G. P. Kuiper, Astrophys. J. **88**, 472 (1938); P. Van De Kamp, Astron. J. **59**, 447 (1954); K. Strand and R. Hall, Astrophys. J. **120**, 322 (1954).

of the same luminosity. According to modern theories of stellar evolution, at least some red giants (those in globular clusters) are much more luminous than main sequence stars of the same mass (Sec. 9). Although "direct" measurements (i.e., analyses of binary systems) of red giant masses are still scanty, we now have at least indirect mass estimates supporting this hypothesis (from theoretical analyses of pulsating red giant stars[4] and from statistics[5] on the angular momentum of main sequence and red giant stars).

Another quantity of interest is the "chemical composition of the universe." We now have rather detailed information on the abundances of the various chemical elements and on isotope ratios in the atmospheres of many stars and in interstellar gas (spectroscopic analysis), as well as in meteorites and even (to some extent) in cosmic rays. Apart from some definite and interesting variations, there is a great similarity between the abundance ratios in such different components of our galaxy. This wealth of data on the "cosmic abundances" of the elements and isotopes[6] has stimulated development of theories on the origin of the elements. For the purpose of calculations on energy production and the interior structure of stars, a knowledge of the abundances of broad *groups* of elements is sufficient. The groups of most interest for this purpose are (together with their approximate *cosmic* abundances in percent, by mass): hydrogen (90%), helium (10%), carbon to neon (1%), and the metal group near iron (0.5%). Although this broad grouping of elements is sufficient, the results of stellar interior calculations are quite sensitive to variations of the chemical composition with radial distance in the interior of a star. Such variations cannot be observed directly and one must investigate the consequences of various assumptions regarding them.

3. STELLAR INTERIOR CALCULATIONS[7]

Calculations on the internal structure of a star in full equilibrium proceed as follows: For a star of a given mass M, *assume* some particular chemical composition as a function of radial distance throughout the star. We do *not* assume any law of energy production at the moment, but merely specify *where* the energy production occurs (whether it all comes from the center of the star or falls off with radial distance in a certain way, etc.). For main sequence stars at least, the energy production is largely concentrated in the central regions of the star and the results are not very sensitive to the exact extent of the energy producing region. The assumption that the star is spherically symmetric and in a steady state leads to two equilibrium equations, one relating the pressure gradient to the gravitational force, the other relating the temperature gradient to the radially outgoing energy flux. These two equations permit a single infinity of solutions for $\rho(r)$, $T(r)$ and $L(r)$—the density, temperature, and energy flux as a function of radial distance r. The single infinity of solutions is connected with the fact that we have not yet specified the law of energy production. If we specify one further quantity, say the central temperature T_c, the solution then gives $\rho(r)$, $T(r)$ and $L(r)$ uniquely and hence also the luminosity $L=L(\infty)$, i.e. the total amount of energy radiated per second. In the more approximate kinds of calculations, ρ and T become zero for a finite value of r and we identify this value with the stellar radius R.

Consider now the case where the rate of energy production (nuclear or otherwise) is known as a function of temperature, density and chemical composition. Since the rate of energy production, integrated over the whole star, must equal the total luminosity L, we can now pick out the one correct solution (i.e., the correct central temperature T_c) out of the singly infinite set (if we have guessed reasonably correctly, in our original assumption, where the bulk of the energy production occurs). In our single infinity of solutions, the luminosity L is fairly insensitive to the value of T_c, but depends quite strongly on the mass and over-all chemical composition of the star. To put it more crudely, the luminosity L is largely determined by the atomic physics and kinetic and gravitation theory which enters the equilibrium equations; the nuclear physics which gives the rate of energy production (which is very temperature sensitive) merely determines the central temperature at which the energy production keeps pace with the luminosity.

4. MAIN SEQUENCE STARS

The simplest assumption about the spatial distribution of chemical composition and energy production (the "stellar model") is the following: We assume *uniform* chemical composition throughout the star and that the energy production is concentrated towards the center of the star. We now have overwhelming evidence to suggest that this kind of stellar model gives a good representation of a main sequence star (at least a fairly young one): We suppose the star was formed from a gas of uniform chemical composition, consisting mainly of hydrogen. The star is in a steady state since its central temperature is just high enough for the slow conversion of hydrogen into helium to produce enough energy to supply the radiation (carbon cycle and proton chain, Sec. 7). The energy production is concentrated towards the center, since the thermonuclear reaction rates are very temperature sensitive and the temperature decreases from the center outwards. Finally, *if*

---

[4] A. R. Sandage, Astrophys. J. **123**, 278 (1956); V. C. Reddish, Monthly Notices Roy Astron. Soc. **115**, 32 (1955).

[5] A. R. Sandage, Astrophys. J. **122**, 263 (1955).

[6] H. S. Brown, Revs. Modern Phys. **21**, 625 (1949); H. E. Suess and H. C. Urey, Revs. Modern Phys. **28**, 53 (1956).

[7] B. Strömgren, Ergeb. exakt. Natur. **16**, 465 (1937); S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (University of Chicago Press, Chicago, 1939); S. Chandrasekhar, *Astrophysics, A Topical Symposium* (McGraw-Hill Book Company, Inc., New York, 1951), Chap. 14; L. H. Aller, *Nuclear Transformations, Stellar Interiors, and Nebulae* (Ronald Press, New York, 1954).

the star is fairly young, the total amount of hydrogen which has been converted into helium in the central regions of the star is small and the chemical composition is still almost uniform.

Mention should be made of some of the sources of inaccuracies in stellar interior calculations:

(i) The rates of nuclear energy production are discussed in Sec. 7. These rates are now known not with very good, but with adequate, accuracy.

(ii) We also need to know the coefficients occurring in one of the two equilibrium relations, the energy transport equation. Throughout most of a star, the energy transport is carried by radiation and the coefficient we need is the opacity. Atomic theory is well enough understood that, in principle, one can calculate the opacity coefficient as a function of temperature, density and chemical composition. In practice such calculations are very tedious but, with the help of computing machines, reliable tables[8] of opacity coefficients have by now been computed.

(iii) On the other hand, we do not yet have sufficiently quantitative theories of turbulence and convection. In the deep interior of some main sequence stars (the hotter ones, which burn hydrogen on the carbon cycle) the energy transport is carried by convection, but in these regions convection is a very efficient process and the calculations are not seriously affected by our poor estimates of convection rates. However, in stars with a low enough surface temperature that the hydrogen in the photosphere is not yet fully ionized, we also have another convection zone just below the photosphere. The fractional mass of the star contained in this outer convection zone is quite small (less than one percent for our sun) and this zone has a negligible effect on the star's interior or on its luminosity. However, we have to contend with the observational definition of the star's "radius" R which, from the theoretical point of view, is rather unfortunate: One observes the radius of the photosphere (about one optical thickness only between it and the outside). which is an extremely tenuous envelope of the star containing only a minute fraction of its mass. The radius of this photosphere is unfortunately quite sensitive to the details of the outer convection zone (and to inaccuracies in the calculation), whereas the root mean square radius for the star's mass, for instance, would be quite insensitive. Nevertheless, quite satisfactory results have been obtained[9] even for main sequence stars cooler than the sun ("red dwarfs"), for which the outer convective zone is particularly important: One assumes a chemical composition of the star (of given mass) equal to that of the star's amosphere One then *adjusts* a parameter, connected with the depth of the hydrogen convection zone, so that the calculated radius agrees with the observed one. The necessary value of this parameter is found to lie within the limits given by the present incomplete theories of convection. One can then calculate the star's luminosity, which is found to be in reasonable agreement with the observed luminosity.

(iv) For main sequence stars of unknown age we have an added uncertainty. In the absence of mixing, whatever helium has been transformed out of hydrogen in the star's past history has accumulated in its deep interior (where the conversion mainly went on) and it now has a slightly nonuniform chemical composition.[10] Calculations for the sun,[11] which neglect the outer convective zone and the chemical inhomogeneity in its interior, have led to discrepancies with observation. However, since the sun's age is known to be 4 or 5 billion years, one can calculate the excess amount of helium in its interior. Calculations,[12] which take account of this chemical inhomogeneity and treat the outer convection zone as described previously, give a luminosity in agreement with the observed one. According to these calculations, the sun is about 30% more luminous now than when it was still young and of homogeneous composition. In fact, the finite width of the main sequence band on a Hertzsprung-Russell diagram for nearby stars of all ages is now thought to be due mainly to the variations in age (and hence in the degree of chemical inhomogeneities): The observed main sequence for a single galactic cluster, consisting of young stars of the same age, is much narrower[13] and one can extrapolate to a well defined "zero-age main sequence" from these observations.

5. EVOLUTIONARY SCHEME

From the equilibrium equations for a steady-state star one can derive a very simple relation somewhat resembling a virial theorem which relates the gravitational to the thermal energy content of the star: For any particular model one can calculate the gravitational potential energy V_{gr} of the star which is of the order of magnitude of $-GM^2/\bar{R}$, where M is the mass and $\bar{R}$ the (root-mean-square) radius of the star. The thermal energy content E_{th} is $Nk\bar{T}$ where N is the total number of ions and $\bar{T}$ is the average temperature which for many models is a factor of 5 or 10 smaller than the central temperature. Our simple relation states that the ratio of $-V_{gr}$ to E_{th} is of the order of and slightly *larger* than unity for most stellar models. Thus in

[8] G. Keller and R. E. Meyerott, Argonne National Laboratory Repts. Nos. 4771 and 4856 (1952) and an unpublished paper; H. Zirin, Astrophys. J. 119, 371 (1954); I. Epstein and L. Motz, Astrophys. J. 120, 156 (1954).

[9] D. E. Osterbrock, Astrophys. J. 118, 529 (1953).

[10] This is particularly serious for stars more luminous than the sun, which burn hydrogen rather quickly. For Sirius A, for instance, see T. L. Swihart, Astrophys. J. **118**, 577 (1953).

[11] I. Epstein and L. Motz, Astrophys. J. **117**, 311 (1953); P. Naur, Astrophys. J. **119**, 365 (1954); G. O. Abell, Astrophys. J. **121**, 430 (1955).

[12] Schwarzschild, Howard, and Härm, Astrophys. J. (to be published).

[13] H. L. Johnson and C. F. Knuckles, Astrophys. J. **122**, 209 (1955); H. L. Johnson and W. A. Hiltner, Astrophys. J. **123**, 267 (1956); Johnson, Sandage, and Wahlquist, Astrophys. J. **124**, 81 (1956).

going from an equilibrium state of one radius to another state of smaller radius the star liberates excess energy[14] which can act as a source for the energy that is radiated away. As a general rule then, when a star has no nuclear energy sources operating in its interior at any stage, the main part of its mass will undergo gravitational contraction. According to our "virial theorem," the mean temperature of the star will also increase during the contraction, in addition to the energy release which feeds the star's luminosity.

Crudely speaking, we can divide the life history of an individual star (according to present-day theories) into three stages: (i) In the first stage the star starts off as a fairly dilute and cold gas mass. Its central temperature is much too low for any thermonuclear reactions to take place. The star is therefore not in a truly steady state, but it will undergo slow gravitational contraction and its luminosity is supplied by the release of excess gravitational energy during the contraction (see Sec. 6). At the same time the temperature in its interior increases towards the values at which nuclear energy production is appreciable. (ii) In the second stage the interior temperatures are just right for the H—He conversion to supply all the energy which goes into the star's luminosity. At this stage the star is on the main sequence, already discussed in Sec. 4, and remains in this steady state for a long time. (iii) After the star has stayed on the main sequence long enough for an appreciable part of its interior to have exhausted its hydrogen content, the star begins to move off the main sequence again. The details of the subsequent evolution are complicated and depend strongly on the star's mass (and possibly on other factors, such as angular momentum). However, at the early phases of this evolution, each star seems to move off the main sequence in the direction of the red giant region, i.e., its luminosity increases and surface temperature decreases. It is now believed that the inner core of the star contracts gravitationally during this early phase, but the radius of its photosphere increases (Sec. 4). At later phases of the evolution, high enough central temperatures may be reached for further nuclear reactions to take place (Secs. 8 and 9).

The comparison of theories on stellar evolution with observational data is helped greatly by two facts: On the observational side we are helped by the existence of various groups of stars for which we have good reason to believe that all the stars inside one group are, at least roughly, of the same age. There are various kinds of such groups: The "globular clusters" each contain a very large number (of the order of 10^5) of stars. These clusters are members of "stellar[15] population II", i.e., stars formed in the early stages of the life of our galaxy as a whole. On the other end of the scale we have the "expanding star associations," containing only a small number (about 100) of stars. Such an association seems to be a dynamic structure and the stars in it were probably formed at the same time. The actual age varies from one association to another, but they are all very young (down to about 10^6 years) and are extreme examples of "stellar population I," Intermediate between these two types of star groups, both regarding age and number of stars, are the "galactic clusters" (or "open clusters").

TABLE I. The time scales t(i), t(ii) and t(iii) for the three evolutionary stages for stars of two different masses M and luminosities L(M_0 and L_0 refer to the sun; the times t are in years).

M	L	t(i)	t(ii)	t(iii)
M_0	L_0	20×10^6	15×10^9	?
$1.3M_0$	$5L_0$	7×10^6	3×10^9	10^8

On the theoretical side we are helped by the fact that an individual star spends a very much longer time "sitting still" on the main sequence (stage ii) than it takes to contract towards the main sequence (stage i) or to evolve away from it (stage iii). The difference between the time-scales of stages (i) and (ii) comes about as follows: The mean thermal energy per ion, $k\bar{T}$, for a main sequence star is less than one kev ($\bar{T}<10^7$°K) and the gravitational energy release per ion during the *whole* gravitational contraction to the main sequence is of the same order of magnitude. On the other hand, the overall energy release in the nuclear H—He conversion is 6.5 Mev per H atom, orders of magnitude more than the gravitational energy content. The time scale of stage (iii) is short partly because the over-all energy release is smaller than in stage (ii) and partly because the luminosity is much larger during stage (iii). Rough figures for the time-scales t(in years) for the three stages are given in Table I for a star of solar mass M_0 and a slightly more massive one.

6. EARLY GRAVITATIONAL CONTRACTION

Let me describe in slightly more detail the first evolutionary stage, during which the star is too cool to burn hydrogen. I will not discuss at all the original "birth of the star," i.e., the processes that lead from an irregular cloud of cold and dilute gas and dust to a spherical mass of gas which (while much cooler than a main sequence star) is hot enough to be optically opaque. These early processes are very complicated, depend strongly on the distribution of dust and turbulence, and we understand the birth process poorly, if at

[14] There are some exceptions to this rule, notably stars of mass less than a certain limit and high enough density that the electrons are degenerate but have nonrelativistic energies. In this case one can get a stable configuration in which no excess energy is liberated either by contracting or expanding the star's radius. White dwarf stars are thought to be in (or near) such configurations.

[15] W. Baade, "Symposium on astrophysics," University of Michigan (July 1953, unpublished report); W. Baade, Scientific Am. 195, 92 (1956); A. R. Sandage, Mem. soc. roy. sci. Liége 14, 254 (1954).

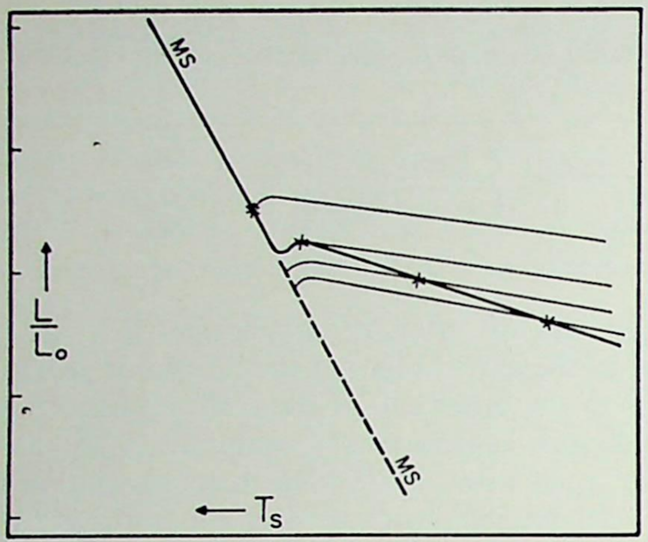


Fig. 2. Schematic Hertzsprung-Russell diagram for a group of young stars of the same age. The heavy solid line is the observed diagram, the dotted heavy line is the missing portion of the main sequence. The light lines are the evolutionary tracks of individual stars, the crosses denote the positions they have reached at the present time.

all. Let us start with a spherical star of low density and central temperature of the order of 10^5°K, so the hydrogen in its interior is hot enough to be ionized but too cold to undergo thermonuclear reactions.

The evolution of such a star can be treated very crudely by means of the "quasi-steady-state approximation": We disregard the question of energy production at the moment, assume a *steady*-state star with uniform chemical composition and find the temperature, density and energy flux throughout the star, as given by the equilibrium equations, for a given initial value of the central temperature T_c. From such a calculation we also obtain the luminosity L and the thermal and gravitational energies E_{th} and V_{gr}. One then repeats this calculation for a series of (successively larger) values of T_c. We then pretend that the contraction and heating up of the actual star is represented by the "movement" along this series of steady-state solutions. The time-scale of this "movement" is obtained by balancing the release of excess gravitational energy against the energy radiated away,

$$\frac{d}{dT_c}[V_{gr}(T_c)+E_{th}(T_c)]\times\frac{dT_c}{dt}=-L(T_c). \qquad (1)$$

We have thus obtained the luminosity and radius of the star as a function of time.[16]

As mentioned before, comparison of evolutionary calculations with observation is helped (i) by the existence of groups of stars of the same age and (ii) by the fact that an individual star spends a much shorter time in its stage of gravitational contraction than it does on the main sequence: The heavy solid curve in Fig. 2 shows (schematcially) the observed Hertzsprung-Russell diagram for a very young star-group[17] (about 10^6 years old). The upper part of the main sequence is present, but the lower part (dotted line, low luminosity stars) is missing. In its place we have low luminosity stars of very large radius (small surface temperature). The light lines represent the evolutionary track of individual stars, the cross on each line represents the stage to which each star has evolved during the common lifetime of the group. The more luminous (more massive) stars evolve more rapidly, have already reached the main sequence and "stay put" on it (at least for times much longer than the age of the group). For the less luminous (and less massive) stars we see a "snapshot picture" of various stages during the gravitational contraction—the earliest stages for the least luminous stars which evolve most slowly. The heavy line ("contraction branch") to the right of the main sequence, the observed "snapshot," does not differ very much from the evolutionary track of an individual star. In other words, stars on different parts of the "contraction branch" have similar masses and will lie close to each other by the time they have all reached the main sequence (stars of *much* smaller mass simply haven't contracted enough to radiate sufficiently in the visible spectral region).

The "quasi-steady-state-approximation" is reasonably good in the present case: The luminosity of a star is given mainly by the equilibrium equations, not by the law of energy production, and is not very different if the energy source is gravitational rather than nuclear. Dynamic effects have been neglected. That is, we have neglected the imbalance between gravitational force and pressure gradient, which produces the inward motion of matter during the contraction. This imbalance is very small and the kinetic energy contained in the contraction itself is very small compared with the thermal or gravitational energy contant. To get a feeling for the orders of magnitude—if there were no pressure gradient at all, the star would collapse in a matter of hours compared with the actual contraction times of the order of 10^6 years. Nevertheless, the quasi-steady approximation is only in semiquantitative agreement with more accurate calculations[18] on the detailed evolution of a star. For instance, the "dip" near the main sequence on the heavy line in Fig. 2, which is found both observationally and in the accurate calculations, would not be obtained from a crude quasi-steady-state calculation.

One minor point bears mentioning: If a star had originally contained large amounts of deuterium, lithium, beryllium, or boron, these substances would have largely "burned up" during the gravitational contraction before reaching the main sequence (see Sec. 7). The Hertzsprung-Russell "snapshot picture" for such a group of stars would then have deviated from that shown in Fig. 2. The absence of clear-cut effects of this nature in the observations to date thus puts an

[16] E. E. Salpeter, Mem. soc. roy. sci. Liége 14, 116 (1954).
[17] M. F. Walker, Astron. J. 59, 333 (1954).
[18] Henye, LeLevier, and Levée, Publ. Astron. Soc. Pacific 67, 154 (1955).

upper limit on the original abundances of these elements. However, these upper limits at present are in any case larger than the various estimates from other considerations for these abundances.

7. THERMONUCLEAR REACTIONS WITH HYDROGEN

From the point of view of stellar energy production, by far the most important thermonuclear reactions are two series of reactions called the proton-proton chain (*p-p*) and carbon-nitrogen (C−N) cycle. The net result of each of these series of reactions is

$$4H^1 \rightarrow He^4 + 2\nu + 26.7 \text{ Mev}. \quad (2)$$

Of this energy release, only about 0.5 Mev (*p-p*) or 1.4 Mev (C−N) are lost in the form of neutrino (ν) kinetic energy, the rest is transformed into thermal energy inside the star. The overwhelming importance as stellar energy sources of these two reaction chains is due the fact that hydrogen constitutes the bulk of stellar matter and that the energy release, per unit mass, is greater in the H−He conversion than for any other nuclear processes in stars. The rate of energy production per g/sec from these two chains is plotted roughly against temperature in Fig. 3. The carbon cycle rate increases much more steeply with temperature and dominates for the more luminous and massive stars. The *p-p* chain dominates slightly for the sun and strongly for all less massive (and less luminous) main sequence stars.

Because of their importance, the proton-chain and carbon cycle have been restudied and reviewed[19] extensively in the last few years so I shall not describe them any further. Instead, I will give some general remarks about the systematics of thermonuclear reactions in stars in general, in particular about those[20] which involve hydrogen as one of the reactants.

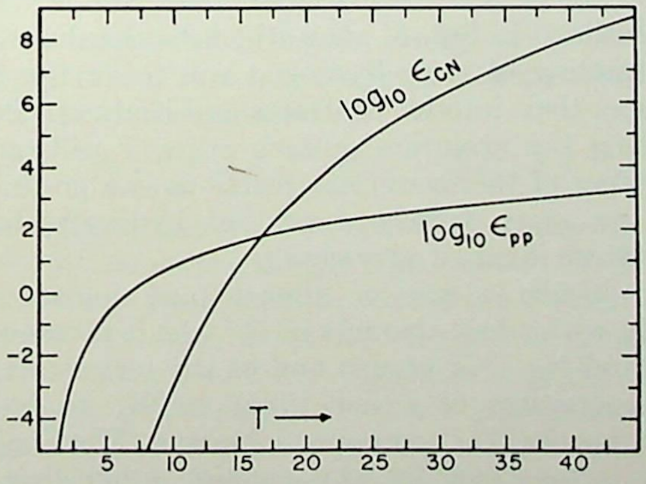


FIG. 3. The energy production on the proton-proton chain (ϵ_{pp}) and on the carbon-nitrogen cycle (ϵ_{CN}), plotted against temperature T (in units of 10^6°K). The curves are for $\rho x_H^2 = 100$ and $x_{CN}/x_H = 0.005$, where ρ is density in g/cc and x_H, x_{CN} are the abundances (by mass) of hydrogen and of the C, N group.

The time scale of these thermonuclear reactions is enormously long, not only compared to nuclear times, but even long compared to reaction times in somewhat related terrestrial devices. The reason for this is essentially the following. The luminosity of the star is essentially given, not by nuclear physics, but just by the opacity and hydrodynamics of the star. Thus, it is really the atomic physics that calls for a particular luminosity, and therefore a particular rate of energy production. That rate is quite small and the total energy content in nuclear reactions is quite large. From these two considerations it follows that the time factor of these thermonuclear reactions in stars has to be of the order of a million to billions of years.

The temperature has to adjust itself in such a way as to give these very long time scales. That is, the temperature must be such that the thermal energy per ion will be extremely small compared to the Coulomb barrier (we will only be talking about reactions between charged particles for the moment). When the ratio between the Coulomb barrier and the thermal energy is of the order of a thousand, the rates are obviously extremely small. One of the consequences of this is that most of the energy production comes from a very narrow energy range, large compared to the thermal energy, small compared to the barrier height, where the product of the Maxwell-Boltzmann factor and the Coulomb barrier penetration factor (both small) has a maximum. For people trying to calculate reaction rates in the stars, this has a rather unfortunate consequence. Since this energy range is rather narrow and also lies at such a low energy, the cross sections cannot be measured directly in the laboratory. Also, they are very sensitive to resonances in the reactions which may occur in or near this narrow energy region and which are very difficult to observe. But at least that keeps both the theorists and experimentalists busy.

You can see then that the rate of these reactions is very temperature-sensitive, and also depends very strongly on the charge of the particles involved. Another thing that follows is that we need only talk of exothermic nuclear reactions, as the endothermic ones are almost impossible when you only have a few kilovolts of kinetic energy. Furthermore, since the time scale is so long, these reactions occur only with stable nuclei; in other words, if a beta-active nucleus is made in one reaction, it will have plenty of time to beta decay before it will absorb another proton or alpha particle.

As to the important reactions where protons are absorbed by various other kinds of nuclei, there are essentially two main trends which determine what is happening. One of them is just that, in the packing

[19] E. E. Salpeter, Phys. Rev. **88**, 547 (1952); E. A. Frieman and L. Motz, Phys. Rev. **89**, 648 (1953); W. A. Fowler, Mem. soc. roy. sci. Liége **14**, 88 (1954); Bosman-Crespin, Fowler, and Humblet, Bull. soc. roy. sci. Liége, Nos. 9–10, 327 (1954).

[20] E. E. Salpeter, Phys. Rev. **97**, 1237 (1955); Fowler, Burbidge, and Burbidge, Astrophys. J. **122**, 271 (1955). The proton-proton reaction does not fit into this scheme, since it involves a beta decay.

fraction curve, the binding energy per nucleon increases up to iron, which implies that in general (p,γ) reactions are exothermic. But that is not the whole story. The main part, or at least the other part, is the fact that alpha particles and also $(N=Z=2n)$ nuclei are particularly stable.[21] For this reason, on $(N=2n+2, Z=2n+1)$ nuclei, (p,α) as well as (p,γ) reactions will in general be exothermic. Even though the (p,γ) reaction would release more energy, a reaction resulting in the emission of a particle has a larger partial width than a radiative transition. This has the result that a (p,α) reaction will usually win out, providing the energy of the alpha particle is not so small that the barrier penetration factor prevents it. This has the important consequence that, given hydrogen and some fairly light nucleus, one does not just get successive building on by (p,γ) reactions and beta decay in one long chain up to the iron group. Rather, one gets cycles of reactions, of which the carbon-nitrogen cycle is by far the most important, but nevertheless only one, example.

The net result in each of these cycles is that of Eq. (2). i.e., the conversion of hydrogen into helium, with the other nuclei acting merely as catalysts: Starting from the $(Z=N=2n)$ nucleus three (p,γ) reactions and two positron-emitting beta decays (occurring in different orders in the different cycles) lead to the $(N=2n+2, Z=2n+1)$ nucleus. The (p,α) reaction on this nucleus then completes the cycle. The competing (p,γ) reaction, if it were important, would convert the original nuclei into the $(Z=N=2n+2)$-nucleus instead of reproducing the lighter nuclei. We give below four examples of such cycles, the four beta-stable nuclei that occur in each and the energy releases of the competing (p,α) and (p,γ) reactions.

(iii) C^{12}, C^{13}, N^{14}, N^{15}; $(p,\alpha)C^{12}+5.0$ Mev;
$(p,\gamma)O^{16}+12.2$ Mev.

(v) Ne^{20}, Ne^{21}, Ne^{22}, Na^{23}; $(p,\alpha)Ne^{20}+2.4$ Mev;
$(p,\gamma)Mg^{24}+11.5$ Mev.

(vi) Mg^{24} Mg^{25}, Mg^{26}, Al^{27}; $(p,\alpha)Mg^{24}+1.6$ Mev; (3)
$(p,\gamma)Si^{28}+11.6$ Mev.

(vii) Si^{28}, Si^{29}, Si^{30}, P^{31}; $(p,\alpha)Si^{28}+1.9$ Mev;
$(p,\gamma)S^{32}+4.6$ Mev.

In both the C—N and Ne—Na cycles, the (p,α) reaction does in fact dominate over the (p,γ) reaction and these nuclei are not lost during their cycles. The branching ratios of the two competing reactions, both in the Mg—Al and the Si—P cycle, depend on the position of resonance levels, which are not yet known, but it is quite likely that the Coulomb barrier for the alpha particle will depress the (p,α) rate below the (p,γ) rate.

In reality, the exceptions to our "cycle rule" are as important as are the good examples: In particular, the cycles with $n=1$, 2, and 4 are missing. For $n=1$, there is no $He^4(p,\gamma)$ reaction and both lithium isotopes are broken down to helium without being replenished. For $n=2$, we "almost" have a Be—B cycle, but the Be^8 formed in the last reaction is unstable and breaks up into two alpha particles. Thus the beryllium and boron isotopes are also broken down into helium without being replenished. Finally, for $n=4$, there is no O—F cycle because (p,α) reactions on O^{17} and O^{18} (as well as on F^{19}) are possible. If the full chain of reactions is uninterrupted, the oxygen and fluorine isotopes are thus converted into isotopes of the carbon-nitrogen group.

To return to the relevance of these proton reactions to stellar interiors: The thermonuclear rates of the various reactions decrease strongly with increasing nuclear charge and increase strongly with temperature. The proton reactions on the Li, Be, and B isotopes are sufficiently fast already at temperatures much smaller than the central temperatures of main sequence stars. Since these nuclei are converted into helium, rather than recycled, they are burned up during the early contraction of a star. In a main sequence star, the bulk of the hydrogen takes part in the C—N cycle (or *p-p* chain). At these temperatures, characteristic for the C—N cycle, the Ne—Na cycle and the $O^{16}(p,\gamma)$ reaction proceed rather slowly. These reactions are unimportant as energy sources but, at least in the hottest stars which still contain an appreciable amount of hydrogen, may be of importance regarding theories of element formation. In particular, these reactions may provide a source of O^{17} and Ne^{21} which, in turn, can undergo reactions which produce neutrons.

8. THERMONUCLEAR REACTIONS WITHOUT HYDROGEN

As outlined in Sec. 5, stars which have exhausted a large fraction of their hydrogen will leave the main sequence; their interior contracts and heats up. Before discussing the evolution of such stars, I will outline what types of thermonuclear reactions can go on in a large gas mass, largely devoid of hydrogen, as its temperature is raised very slowly.

We consider a gas of almost pure ionized He^4, possibly with small amounts of O^{16}, the isotopes of the C—N and Ne—Na groups and of the metals present. At temperatures of a few times 10^7°K, no nuclear reactions proceed at any appreciable rate: The Coulomb barrier is too large for any reaction other than the collision of two alpha particles, but such a collision cannot lead to any exothermic reaction (Be^8 is unstable). The only exothermic reaction which can occur in pure helium is the conversion of three alpha particles into one carbon nucleus,[22]

$$3\,He^4 \rightarrow C^{12}+\gamma+7.66 \text{ Mev}. \quad (4)$$

[21] N and Z are the number of neutrons and protons in the nucleus, n is an integer. We are considering fairly light nuclei at the moment.

[22] E. J. Öpik, Proc. Roy. Irish Acad. **A54**, 49 (1951); E. E. Salpeter, Astrophys. J. **115**, 326 (1952).

The reaction rate for such a "triple collision" might be expected to be extremely slow at densities of 10^3 to 10^5 g/cc, unless the temperature is very high indeed. In fact, the speed of the reaction in Eq. (4) is increased enormously by the chance coincidence of two resonance levels in two nuclei occurring in the most opportune energy regions: Although Be^8 is not stable, its ground-state energy is not much higher than that of two alpha particles with thermal kinetic energies (for temperatures of about 10^8°K). We thus have a dynamic equilibrium set up in which a certain small fraction of the matter is in the form of Be^8. The other resonance of interest[23] is the second excited state in C^{12}, which lies at an energy slightly larger than that of a Be^8 nucleus plus an alpha particle plus thermal kinetic energy. The He−C conversion thus takes place by way of an (α,γ) reaction on Be^8 through the second excited level of C^{12}.

Much experimental and theoretical work has been done on this resonance level in C^{12}, the most critical part of the key reaction to all later reactions. The rates of this reaction, at various temperatures and densities, are known to within about a factor of ten either way[24]: The energy of the C^{12} level has been measured very accurately and it has been verified that it has the right spin and parity to be formed from Be^8 plus an alpha particle. The same experiments have given an upper limit for the partial width of transitions to the ground state from this level. Inelastic electron scattering experiments at Stanford give, indirectly, a lower limit for this width and theoretical estimates for its actual value have also been made.

The $3\alpha \rightarrow C^{12}$ reaction time is of the right order of magnitude (10^6 to 10^7 years) for stellar energy production at temperatures of about 1×10^8°K (for densities of the order of 10^4 g/cc). At these temperatures, further (α,γ) reactions also go on, starting with the C^{12} which has been formed. A chain of such (α,γ) reactions could, in principle, build up $(N=Z=2n)$ nuclei up to Ca^{40}. At temperatures of about 1×10^8°K, however, only the reactions

$$C^{12}(\alpha,\gamma)O^{16}+7.2\text{ Mev};\quad O^{16}(\alpha,\gamma)Ne^{20}+4.8\text{ Mev} \tag{5}$$

proceed rapidly enough. The net effect of these reactions is then the depletion of He^4, the buildup of C^{12}, O^{16}, and Ne^{20} (and, possibly, a little Mg^{24}), accompanied by an energy release of about 0.9 Mev per nucleon (compared with about 6.5 Mev per nucleon in the H→He conversion). The reaction rates are known accurately enough for the purpose of stellar energy production, but are not yet accurate enough from the point of view of element-synthesis: In the $O^{16}(\alpha,\gamma)$ reaction, in particular, the possible effects of two resonance levels in the right energy region are still unknown. We therefore do not know yet the abundance ratio of O^{16} to Ne^{20}, which will be produced by this chain of reactions.

Another type of reaction is of importance for element-synthesis at these (or even slightly lower) temperatures. Again because of the great stability of light $(Z=N=2n)$ nuclei, (α,n) reactions on some $(Z=2n, N=2n+1)$ nuclei are exothermic. In particular, (α,n) reactions on the nuclei C^{13}, O^{17}, and Ne^{21} proceed rapidly enough[25] at these temperatures. These isotopes are much too rare for reactions involving them to serve as a significant energy source. On the other hand, even fairly small amounts of neutrons may be of importance in synthesizing many of the less abundant isotopes. In addition, the nuclei C^{13}, O^{17}, and Ne^{21} may be replenished to some extent (Sec. 7) if small amounts of hydrogen are introduced into the gas mass (e.g., by mixing in from the outer layers of the star which are still hydrogen-rich).

Calculations have also been made[26] on reactions that might take place at still higher temperatures in a gas which is helium-deficient, as well as containing no hydrogen, i.e., consists mainly of C^{12}, O^{16}, and Ne^{20}. At temperatures of about 1×10^9°K two different types of reactions become important. The one type involves the collision between two of the more abundant nuclei. The reaction proceeding at the lowest temperature (from about 0.5×10^9°K up) involves the collision of two C^{12} nuclei, resulting in one of the following three possibilities

$$Mg^{24}+\gamma+14\text{ Mev},\quad Ne^{20}+\alpha+5\text{ Mev},\quad Na^{23}+p+3\text{ Mev}. \tag{6}$$

The other types of reaction involve first of all endothermic (γ,α) reactions. These are initiated by photons with energies "way up in the tail" of the blackbody radiation spectrum. These photodisintegrations are rare, even at these high temperatures, but are followed immediately by the inverse processes, (α,γ) captures, on various nuclei. The net result of all these reactions is a tendency for the lighter and less stable nuclei to be destroyed and the slightly heavier and more stable nuclei, up to about silicon or sulfur $(A\sim30)$, to be built up. At slightly higher temperatures still, about $(2\text{ to }3)\times10^9$°K, a whole host of reactions begins to take place. Besides the types of reactions already mentioned, (γ,p) and (p,γ) reactions followed by beta decays are also possible, which can result in nuclei containing more neutrons than protons. The end result of all these reactions, at about 3×10^9°K, is a tendency toward statistical equilibrium of the relative abundances of the nuclides according to their binding energy per nucleon; i.e., a peak is built up in the region of the most stable nuclei, in the neighborhood of Fe^{56}.

---

[23] Dunbar, Pixley, Wengel, and Whaling Phys. Rev. **92**, 649 (1953).

[24] (a) E. E. Salpeter in "Symposium on astrophysics," University of Michigan (July 1953, unpublished report); (b) F. Hoyle, Astrophys. J. Suppl. Ser. 1, 121 (1954); (c) Hayakawa, Hayashi, Imoto, and Kikuchi, Progr. Theoret. Phys. (to be published); (d) W. A. Fowler *et al.* Phys. Rev. (to be published).

[25] A. G. Cameron, Astrophys. J. **121**, 144 (1955); Fowler, Burbidge, and Burbidge, Astrophys. J. **122**, 271 (1955).

[26] See references 24(a) and (b). S. Nakamura (unpublished work).

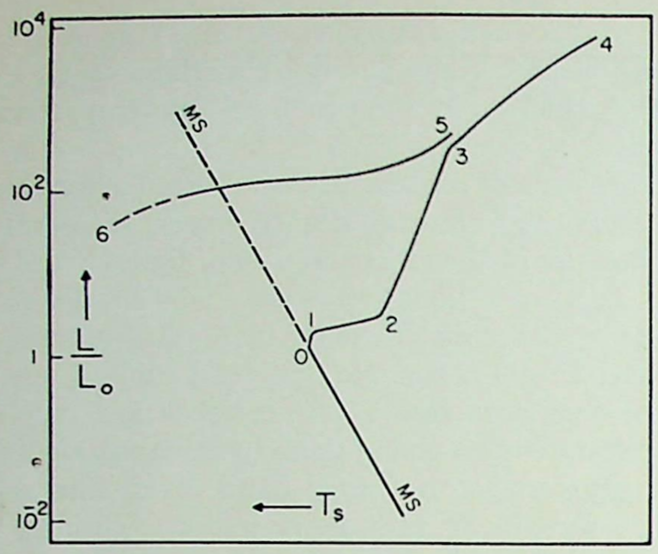


FIG. 4. Schematic Hertzsprung-Russell diagram for a typical globular cluster. The dotted line represents the missing portion of the main sequence.

The energy release per mass in all these later reactions is quite small: The whole chain of reactions leading from O^{16} to Fe^{56}, say, only produce a total energy release of about 0.6 Mev per nucleon. To summarize, we have found possible nuclear energy sources at three different temperature ranges: (i) The H→He conversion at (1 to 3)$\times 10^7$°K with a large release of energy. (ii) The He→C, O, Ne conversion at (1 to 2)$\times 10^8$°K with a smaller energy release. (iii) A host of reactions converting the last group into the iron-group at (0.5 to 3)$\times 10^9$°K with a very small energy release.

9. EVOLUTION OF POPULATION II RED GIANTS

Let us return now to stage (iii), outlined in Sec. 5, in the evolution of an individual star, its history after it leaves the main sequence. We now have a large amount of observational data on the Hertzsprung-Russell diagram for stars[27] in globular clusters (extreme cases of stellar population II). The various globular clusters (seven of them have been analyzed in detail so far) are thought to be of about the same age (5 to 6×10^9 years) and have remarkably similar Hertzsprung-Russell diagrams, shown schematically in Fig. 4.

As Fig. 4 shows, the upper part of the main sequence is missing, the break (point 0) occurring at stars of about twice the sun's luminosity (mass about 1.1 M_0). Instead, we have the curve "to the right and up" of the main sequence, with fairly definite changes of slope at the points marked 1 to 6. According to calculations described below, the evolution of an individual star off the main sequence is very rapid (especially after point 2). The "snapshot picture" in Fig. 4 lies then fairly close to the evolutionary track traced out by an individual star and the masses of the various stars along various parts of the snapshot vary very little (from about $1.1M_0$ at point 1 to about 1.2_5M_0 at point 6).

Much theoretical work has been done in recent years on the internal structure of such stars. The latest calculations[28] are in remarkably good semiquantitative agreement with the observational data. I shall only outline the physical features implied by the theoretical results: Until the star has converted about 10 to 15% of its hydrogen mass into helium (point 1) it does not move far off the main sequence (our sun has progressed about halfway towards point 1). But, by this stage, it has developed a marked inhomogeneity: We have a central core, consisting mainly of inert helium at a uniform temperature, surrounded by a hydrogen-rich spherical shell in which the H−He energy production now takes place. As more hydrogen "burns out," the fraction q of the star's mass inside the helium core grows and the structure of the star varies quite markedly with q:

As the core mass q increases, the density of the core increases, the outer envelope of the star becomes more tenuous (radius increases) and the luminosity increases. The time scale of this evolution is determined by the rate at which hydrogen must be burned up (adding its mass to q) in order to supply the luminosity. Between point 1 and about point 3, this rate is small enough so that the rate of release of gravitational energy during the contraction of the core is small. The core is then almost isothermal, does not heat up much while its density increases (the temperature of the hydrogen-burning shell is kept near 3×10^7°K by the carbon cycle) and the electrons in the core's interior are degenerate. Near point 3, however,[29] the rate of the core contraction increases. The outward flow of the released gravitational energy produces a temperature gradient and the central temperature increases during the speedup until it reaches about 1.2×10^8°K at point 4 (the central density is now more than 10^5 g/sec). At this stage the energy production in the center of the core due to the $3\alpha\rightarrow$C conversion (Sec. 8) becomes important. This central energy production, in a region in which the electrons are degenerate, now leads to a very fast expansion of the core which continues until the electrons are again nondegenerate (density less than 10^4 g/sec). This is reflected in a fairly sudden jump from point 4 to point 5 in Fig. 4 (the core mass q is about 0.5 at this stage). Finally, in its evolution from point 5 toward point 6, the star has two steady nuclear energy sources, the $3\alpha\rightarrow$C conversion near the center of the core and the H−He conversion in the shell outside the core.

The calculations, whose results I have just described, use a number of rather drastic simplifying approximations. The numerical results are thus not yet very accurate, but the semiquantitative agreement with observation is already quite remarkable. Another

[27] Arp, Baum, and Sandage, Astron. J. 58, 4 (1953); A. R. Sandage, Mem. roy. sci. Liége 14, 254 (1954); H. C. Arp, Astron. J. 60, 317 (1955).

[28] A. R. Sandage and M. Schwarzschild, Astrophys. J. 116, 463 (1952); F. Hoyle and M. Schwarzschild, Astrophys. J. Suppl. Ser., 2, 1 (1955).

[29] The change in slope at point 2 is due to the onset of convection in the outermost envelope of the star.

word of warning: Even the qualitative features of the evolutionary track of an individual star, after it leaves the main sequence, are quite sensitive to the star's mass and the development of a star of more than twice the solar mass is probably rather different[30] from that of globular cluster stars. Nevertheless, we have both observational and theoretical qualitative data which indicate that younger and more massive stars in population *I* also evolve away from the main sequence towards larger luminosities and radii (red giant region). We cannot, of course, have any direct observational evidence for the presence of the nuclear reactions, described in Sec. 8, in the interior of stars. Nevertheless, anomalous abundances of various elements and isotopes in the surface of certain types of red giant stars at least hint at the formation of C^{12} to Ne^{20} (and possibly also of some of the heavier elements).

10. STELLAR EXPLOSIONS

I would like to conclude with two speculative ideas on the possibilities of stellar explosions; i.e., thermonuclear reactions which do not proceed on a long time scale, but on a very short one. The first one has to do with the expanding associations of gas and dust, mentioned in Sec. 5. Such an association seems to be a dynamic unit, containing a mass of a few hundred M_0 or more, and has quite a large amount of kinetic energy stored in its high expansion velocity. The problem is, what was the original energy source which fed this enormous expansion?

The most spectacular of various attempts at an explanation of this phenomenon is the following.[31] Assume that the original gas mass contracted under very special circumstances, for instance, by having very little angular momentum initially and by having density fluctuations in just the right way so that the whole thing fell in at a rapid rate. It heated up as it contracted, but had enough inwardly directed momentum so that it overshot the point where the first thermonuclear reaction would ordinarily stop the contraction. An appreciable fraction of the hydrogen in the central regions of this contracting gas mass might thus be converted into helium very suddenly. The enormous amount of energy produced raises the central pressure, which finally overcomes the inwardly directed momentum and an explosion (fast expansion) results. Since the exploding region is surrounded by a gas mass of very large radius, the original explosion would probably not be seen in terms of visible light. This explosion would eventually lead to a spherical expanding shock wave and bright stars might be formed out of the gas in a region near the shock front.[32]

The above hypothesis is, at the moment, highly

[30] R. J. Tayler, Astrophys. J. 120, 332 (1954).
[31] Unpublished work by J. F. Bird (Cornell University), following a suggestion by T. Gold.
[32] J. H. Oort and L. Spitzer, Astrophys. J. 121, 6 (1955); M. P. Savedoff and J. Greene, Astrophys. J. 122, 477 (1955).

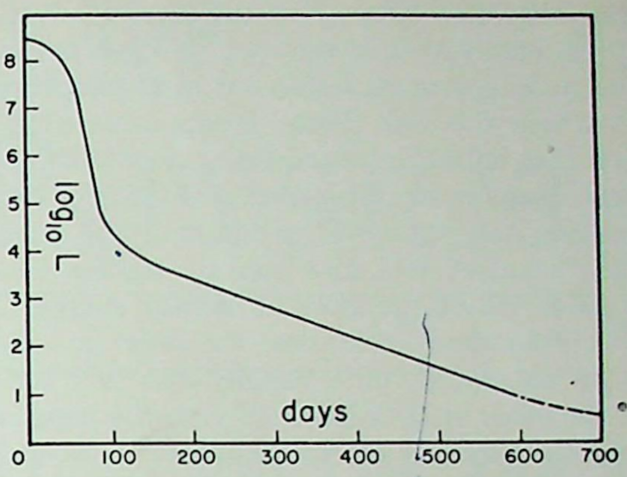


Fig. 5. The (schematic) light curve of a supernova of Type I. *L* is the luminosity in arbitrary units, plotted against time (in days after the initial eruption).

speculative, but it also serves as an introduction to one part of the other hypothesis I want to talk about. This has to do with the 55-day half-life of the light curve of supernovae of Type I and with the fission decay of Cf^{254} with the same half-life. In Fig. 5, the life history of the logarithm of the luminosity of a supernova, after it has first erupted, is shown. There is an enormous peak for the first 100 days or so, and then there is a very peculiar "tail" which, total energy-wise, is only about 10^{-3} of the original hump. This tail is an exponential curve for about 600 days with a 55-day half-life. The same half-life is found among all supernovae of this type (Type I). One has little direct observational evidence for what happens afterwards but, as you will hear from Professor Morrison in his talk on the origins of cosmic rays, there is at least indirect evidence that 1000 years later, in the Crab nebulae, there still is something or other that gives off energy. Thus, the contribution of the "Cf^{254} decay" is just a tiny part in quite a long and complicated light curve. It is, nevertheless, very spectacular to find a pure exponential decay for a period of over 10 times the half-life.

Previous theories on supernovae all had the one essential thing in common. One starts with the core of the star composed to a large extent of the most stable nuclei, the iron group. Then it is argued that, since this is the most stable possibility, any further nuclear reactions can only absorb energy. This, however, can lead to a collapse in the following way. If there is a nuclear process which is not only endothermic, but whose rate increases with density, then this will certainly result in collapse. First there is some energy loss, then the star contracts a bit, then its density goes up and there is a higher rate for the reaction, and so on. There are various kinds of such theories, but that is the essence of them. The one that is being investigated the most and works out the best is Gamow and Schönberg's[33] Urca process.

[33] G. Gamow and M. Schönberg, Phys. Rev. 59, 539 (1941); also G. Gamow (unpublished work).

It was fairly easy to see how the core of a star could collapse suddenly. What was not so easy to see was how one could get an explosion out of a collapsing star. This is where the Cal. Tech. group came in.[34] They observed that only the core of the star need be entirely composed of the most stable nuclei. Outside of the core, say the outer half, there will undoubtedly be some iron, but this "mantle" will also contain lighter elements with a potentially large store of nuclear energy. Then, as the core suddenly collapses, the outer mantle will also begin to shrink quite rapidly and will heat up. Like in the other story I just told you, the hope is that the inwardly directed momentum during the collapse of this mantle will be large enough so that one gets a large rate of thermonuclear reactions before the contraction is brought to a halt. The star will then blow up and, if the mechanism was right, it will have put quite a bit of energy into the expanding cloud.

A much more specific part of the theory depends on the details of the chemical composition of this mantle and must explain the observed 55 day half-life. Briefly, the Cal. Tech. group's theory assumes that there occur roughly equal proportions of hydrogen, helium, and the elements from C to Ne. Given this initial chemical composition and with the right cross-section values (which we are not quite sure about yet), neutrons are made in essentially the following way. First Ne^{21} is made by a (p,γ) reaction on Ne^{20}, and then neutrons come off from just the (α,n) reactions already discussed (Sec. 8) as going on in "ordinary" stars at about 10^{8}°K. This time, however, the reactions proceed very, very rapidly. Once neutrons are made, they can be captured by nuclei in the iron group. It is possible to build all the way up to nuclei with $A\sim 270$ by alternating β decays with the neutron captures. When the whole star explodes, at least some of the products will have alpha decayed down to Cf^{254} which has a fission life of just the right time and a very large energy release per particle. Many quantitative aspects of the "Cal. Tech. theory" still remain to be tested, but it already seems like a very promising idea.

[34] Burbidge, Hoyle, Burbidge, Christy, and Fowler, Phys. Rev. **103**, 1145 (1956).

Development Fund of the American Institute of Physics

Everyone whose career depends on, or has been helped by, the science of physics is invited to contribute to the Development Fund of the American Institute of Physics. The fund will be used to provide larger working space for the Institute's publishing and other services, to launch an expansion of periodical media for the publication of research reports and to encourage and improve the study of physics at elementary levels.

It is hoped to raise $150,000 from individuals and $350,000 from industry. Half of the total will be used to convert, for office use, a building on East 45th Street near United Nations Plaza. The building has been acquired with the proceeds of the sale of the old building at 57 East 55th Street. Removal to the new and much larger headquarters is scheduled for June, 1957.

With the enlarged space of the new building, it will be possible to continue furnishing the physics journals at a subscription price per word considerably lower than is commercially feasible. With expanded journals, more complete reporting of physics research can be provided.

The AIP now publishes over 19 000 pages of journals per year, serving over 17 000 members and a total of more than 69 000 subscribers. A high standard of economy of publishing is maintained. The journals bear no purchase cost or depreciation charge for the equipment or building space used for them by the Institute. Aside from publishing, the AIP helps to exert the concerted effort of the national societies of physics to achieve recognition for physics, to improve teaching, to expand opportunities and facilities for research, and to cooperate with other societies in advancing science and technology in America.

The help of all friends of physics is urgently needed. Checks should be made payable, and sent, to "AIP DEVELOPMENT FUND," 57 East 55th Street, New York 22, New York.

REVIEWS OF

MODERN PHYSICS

VOLUME 29, NUMBER 3 JULY, 1957

Relativistic Invariance and Quantum Phenomena*

EUGENE P. WIGNER

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

INTRODUCTION

THE principal theme of this discourse is the great difference between the relation of special relativity and quantum theory on the one hand, and general relativity and quantum theory on the other. Most of the conclusions which will be reported on in connection with the general theory have been arrived at in collaboration with Dr. H. Salecker,[1] who has spent a year in Princeton to investigate this question.

The difference between the two relations is, briefly, that while there are no conceptual problems to separate the theory of special relativity from quantum theory, there is hardly any common ground between the general theory of relativity and quantum mechanics. The statement, that there are no conceptual conflicts between quantum mechanics and the special theory, should not mean that the mathematical formulations of the two theories naturally mesh. This is not the case, and it required the very ingenious work of Tomonaga, Schwinger, Feynman, and Dyson[2] to adjust quantum mechanics to the postulates of the special theory and this was so far successful only on the working level. What is meant is, rather, that the concepts which are used in quantum mechanics, measurements of positions, momenta, and the like, are the same concepts in terms of which the special relativistic postulate is formulated. Hence, it is at least possible to formulate the requirement of special relativistic invariance for quantum theories and to ascertain whether these requirements are met. The fact that the answer is more nearly *no* than *yes*, that quantum mechanics has not yet been fully adjusted to the postulates of the special theory, is perhaps irritating. It does not alter the fact that the question of the consistency of the two theories can at least be formulated, that the question of the special relativistic invariance of quantum mechanics by now has more nearly the aspect of a puzzle than that of a problem.

This is not so with the general theory of relativity. The basic premise of this theory is that coordinates are only auxiliary quantities which can be given arbitrary values for every event. Hence, the measurement of position, that is, of the space coordinates, is certainly not a significant measurement if the postulates of the general theory are adopted: the coordinates can be given any value one wants. The same holds for momenta. Most of us have struggled with the problem of how, under these premises, the general theory of relativity can make meaningful statements and predictions at all. Evidently, the usual statements about future positions of particles, as specified by their coordinates, are not meaningful statements in general relativity. This is a point which cannot be emphasized strongly enough and is the basis of a much deeper dilemma than the more technical question of the Lorentz invariance of the quantum field equations. It pervades all the general theory, and to some degree we mislead both our students and ourselves when we calculate, for instance, the mercury perihelion motion without explaining how our coordinate system is fixed in space, what defines it in such a way that it cannot be rotated, by a few seconds a year, to follow the perihelion's apparent motion. Surely the x axis of our coordinate system could be defined in such a way that it pass through all successive perihelions. There must be some assumption on the nature of the coordinate system which keeps it from following the perihelion. This is not difficult to exhibit in the case of the motion of the perihelion, and it would be useful to exhibit it. Neither is this, in general, an academic point, even

* Address of retiring president of the American Physical Society, January 31, 1957.

[1] This will be reported jointly with H. Salecker in more detail in another journal.

[2] See, e.g., J. M. Jauch and F. Rohrlich, *The Theory of Protons and Electrons* (Addison-Wesley Press, Cambridge, Massachusetts, 1955).

though it may be academic in the case of the mercury perihelion. A difference in the tacit assumptions which fix the coordinate system is increasingly recognized to be at the bottom of many conflicting results arrived at in calculations based on the general theory of relativity. Expressing our results in terms of the values of coordinates became a habit with us to such a degree that we adhere to this habit also in general relativity where values of coordinates are not *per se* meaningful. In order to make them meaningful, the mollusk-like coordinate system must be somehow anchored to space-time events and this anchoring is often done with little explicitness. If we want to put general relativity on speaking terms with quantum mechanics, our first task has to be to bring the statements of the general theory of relativity into such form that they conform with the basic principles of the general relativity theory itself. It will be shown below how this may be attempted.

RELATIVISTIC QUANTUM THEORY OF ELEMENTARY SYSTEMS

The relation between special theory and quantum mechanics is most simple for single particles. The equations and properties of these, in the absence of interactions, can be deduced already from relativistic invariance. Two cases have to be distinguished: the particle either can, or cannot, be transformed to rest. If it can, it will behave, in that coordinate system, as any other particle, such as an atom. It will have an intrinsic angular momentum called J in the case of atoms and spin S in the case of elementary particles. This leads to the various possibilities with which we are familiar from spectroscopy, that is spins 0, $\frac{1}{2}$, 1, $\frac{3}{2}$, 2, $\cdots$ each corresponding to a type of particle. If the particle cannot be transformed to rest, its velocity must always be equal to the velocity of light. Every other velocity can be transformed to rest. The rest-mass of these particles is zero because a nonzero rest-mass would entail an infinite energy if moving with light velocity.

Particles with zero rest-mass have only two directions of polarization, no matter how large their spin is. This contrasts with the $2S+1$ directions of polarization for particles with nonzero rest-mass and spin S. Electromagnetic radiation, that is, light, is the most familiar example for this phenomenon. The "spin" of light is 1, but it has only two directions of polarization, instead of $2S+1=3$. The number of polarizations seems to jump discontinuously to two when the rest-mass decreases and reaches the value 0. Bass and Schrödinger[3] followed this out in detail for electromagnetic radiation, that is, for $S=1$. It is good to realize, however, that this decrease in the number of possible polarizations is purely a property of the Lorentz transformation and holds for any value of the spin.

There is nothing fundamentally new that can be said about the number of polarizations of a particle and the principal purpose of the following paragraphs is to illuminate it from a different point of view.[4] Instead of the question: "Why do particles with zero rest-mass have only two directions of polarization?" the slightly different question, "Why do particles with a finite rest-mass have more than two directions of polarization?" is proposed.

The intrinsic angular momentum of a particle with zero rest-mass is parallel to its direction of motion, that is, parallel to its velocity. Thus, if we connect any internal motion with the spin, this is perpendicular to the velocity. In case of light, we speak of transverse polarization. Furthermore, and this is the salient point, the statement that the spin is parallel to the velocity is a relativistically invariant statement: it holds as well if the particle is viewed from a moving coordinate system. If the problem of polarization is regarded from this point of view, it results in the question, "Why can't the angular momentum of a particle with finite rest-mass be parallel to its velocity?" or "Why can't a plane wave represent transverse polarization unless it propagates with light velocity?" The answer is that the angular momentum *can* very well be parallel to the direction of motion and the wave *can* have transverse polarization, but these are not Lorentz invariant statements. In other words, even if velocity and spin are parallel in one coordinate system, they do not appear to be parallel in other coordinate systems. This is most evident if, in this other coordinate system, the particle is at rest: in this coordinate system the

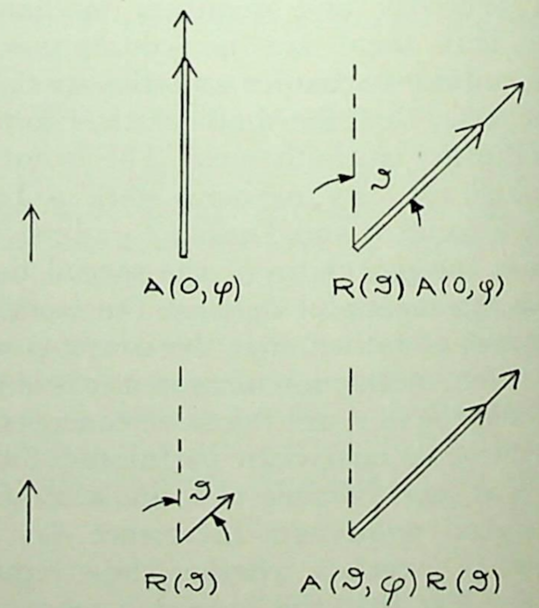


FIG. 1. The short simple arrows illustrate the spin, the double arrows the velocity of the particle. One obtains the same state, no matter whether one first imparts to it a velocity in the direction of the spin, then rotates it $(R(\vartheta)A(0,\varphi))$, or whether one first rotates it, then gives a velocity in the direction of the spin $(A(\vartheta,\varphi)R(\vartheta))$. See Eq. (1.3).

[3] L. Bass and E. Schrödinger, Proc. Roy. Soc. (London) **A232**, 1 (1955).

[4] The essential point of the argument which follows is contained in the present writer's paper, Ann. Math. **40**, 149 (1939) and more explicitly in his address at the Jubilee of Relativity Theory, Bern, 1955 (Birkhauser Verlag, Basel, 1956), A. Mercier and M. Kervaire, editors, p. 210.

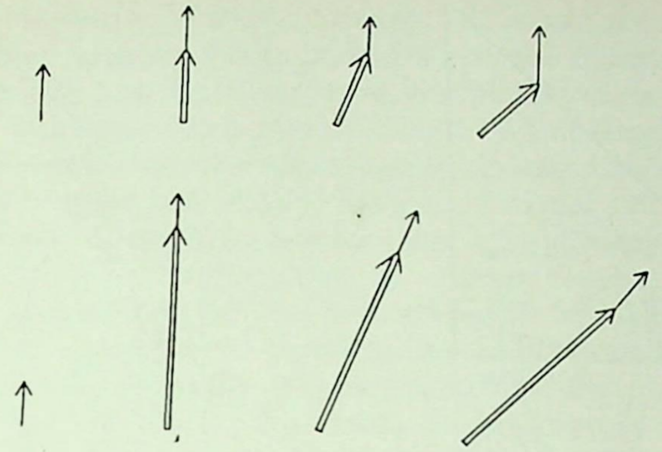


FIG. 2. The particle is first given a small velocity in the direction of its spin, then increasing velocities in a prependicular direction (upper part of the figure). The direction of the spin remains essentially unchanged; it includes an increasingly large angle with the velocity as the velocity in the perpendicular direction increases. If the velocity imparted to the particle is large (lower part of the figure), the direction of the spin seems to follow the direction of the velocity. See Eqs. (1.8) and (1.7).

angular momentum should be parallel to nothing. However, every particle, unless it moves with light velocity, can be viewed from a coordinate system in which it is at rest. In this coordinate system its angular momentum is surely not parallel to its velocity. Hence, the statement that spin and velocity are parallel cannot be universally valid for the particle with finite rest-mass and such a particle must have other states of polarization also.

It may be worthwhile to illustrate this point somewhat more in detail. Let us consider a particle at rest with a given direction of polarization, say the direction of the z axis. Let us consider this particle now from a coordinate system which is moving in the $-z$ direction. The particle will then appear to have a velocity in the z direction and its polarization will be parallel to its velocity (Fig. 1). It will now be shown that this last statement is nearly invariant if the velocity is high. It is evident that the statement is entirely invariant with respect to rotations and with respect to a further increase of the velocity in the z direction. This is illustrated at the bottom of the figure. The coordinate system is first turned to the left and then given a velocity in the direction opposite to the old z axis. The state of the system appears to be exactly the same as if the coordinate system had been first given a velocity in the $-z$ direction and then turned, which is the operation illustrated at the top of the figure. The state of the system appears to be the same not for any physical reason but because the two coordinate systems are identical and they view the same particle (see Appendix I).

Let us now take our particle with a high velocity in the z direction and view it from a coordinate system which moves in the $-y$ direction. The particle now will appear to have a momentum also in the y direction, its velocity will have a direction between the y and z axes (Fig. 2). Its spin, however, will not be in the direction of its motion any more. In the nonrelativistic case, that is, if all velocities are small as compared with the velocity of light, the spin will still be parallel to z and it will, therefore, enclose an angle with the particle's direction of motion. This shows that the statement that the spin is parallel to the direction of motion is not invariant in the nonrelativisitic region. However, if the original velocity of the particle is close to the light velocity, the Lorentz contraction works out in such a way that the angle between spin and velocity is given by

$$\tan (\text{angle between spin and velocity}) = (1-v^2/c^2)^{\frac{1}{2}} \sin\vartheta, \quad (1)$$

where ϑ is the angle between the velocity v in the moving coordinate system and the velocity in the coordinate system at rest. This last situation is illustrated at the bottom of the figure. If the velocity of the particle is small as compared with the velocity of light, the direction of the spin remains fixed and is the same in the moving coordinate system as in the coordinate system at rest. On the other hand, if the particle's velocity is close to light velocity, the velocity carries the spin with itself and the angle between direction of motion and spin direction becomes very small in the moving coordinate system. Finally, if the particle has light velocity, the statement "spin and velocity are parallel" remains true in every coordinate system. Again, this is not a consequence of any physical property of the spin, but is a consequence of the properties of Lorentz transformations: it is a kind of Lorentz contraction. It is the reason for the different behavior of particles with finite, and particles with zero, rest-mass, as far as the number of states of polarization is concerned. (Details of the calculation are in Appendix I.)

The preceding consideration proves more than was intended: it shows that the statement "spin and velocity are parallel for zero mass particles" is invariant and that, for relativistic reasons, one needs only *one* state of polarization, rather than *two*. This is true as far as proper Lorentz transformations are concerned. The second state of polarization, in which spin and velocity are antiparallel, is a result of the reflection symmetry. Again, this can be illustrated on the example of light: right circularly polarized light appears as right circularly polarized light in all Lorentz frames of reference which can be continuously transformed into each other. Only if one looks at the right circularly polarized light in a mirror does it appear as left circularly polarized light. The postulate of reflection symmetry allows us to infer the existence of left circularly polarized light from the existence of right circularly polarized light—if there were no such reflection symmetry in the real world, the existence of *two* modes of polarization of light, with virtually identical properties, would appear to be a miracle. The situation is entirely different for particles with

nonzero mass. For these, the $2S+1$ directions of polarization follow from the invariance of the theory with respect to proper Lorentz transformations. In particular, if the particle is at rest, the spin will have different orientations with respect to coordinate systems which have different orientations in space. Thus, the existence of all the states of polarization follow from the existence of one, if only the theory is invariant with respect to proper Lorentz transformations. For particles with zero rest-mass, there are only two states of polarization, and even the existence of the second one can be inferred only on the basis of reflection symmetry.

REFLECTION SYMMETRY

The problem and existence of reflection symmetry have been furthered in a brilliant way by recent theoretical and experimental research. There is nothing essential that can be added at present to the remarks and conjectures of Lee, Yang, and Oehme, and all that follows has been said, or at least implied, by Salam, Lee, Yang, and Oehme.[5] The sharpness of the break with past concepts is perhaps best illustrated by the cobalt experiment of Wu, Ambler, Hayward, Hoppes, and Hudson.

The ring current—this may be a permanent current in a superconductor—creates a magnetic field. The Co source is in the plane of the current and emits β particles (Fig. 3). The whole experimental arrangement, as shown in Fig. 3, has a symmetry plane and, if the principle of sufficient cause is valid, the symmetry plane should remain valid throughout the further fate of the system. In other words, since the right and left sides of the plane had originally identical properties, there is no sufficient reason for any difference in their properties at a later time. Nevertheless, the intensity of the β radiation is larger on one side of the plane than the other side. The situation is paradoxical no matter what the mechanism of the effect is—in fact, it is most paradoxical if one disregards its mechanism and theory entirely. If the experimental circumstances can be idealized as indicated, even the principle of sufficient cause seems to be violated.

It is natural to look for an interpretation of the experiment which avoids this very far-reaching conclusion and, indeed, there is such an interpretation.[5a] It is good to reiterate, however, that no matter what interpretation is adopted, we have to admit that the symmetry of the real world is smaller than we had thought. However, the symmetry may still include reflections.

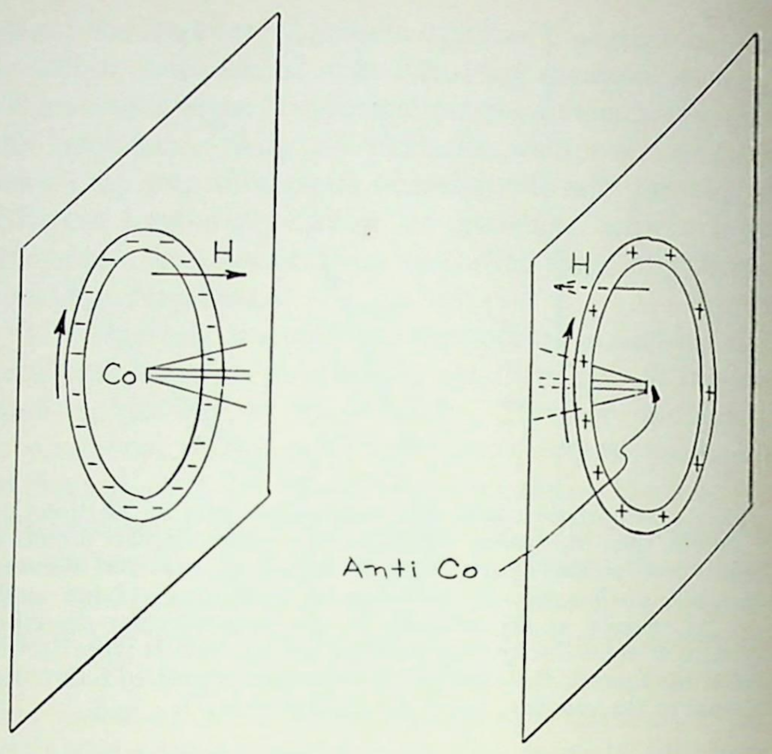


FIG. 3. The right side is the mirror image of the left side, according to the interpretation of the parity experiments[5a] which maintains the reflection as a symmetry element of all physical laws. It must be assumed that the reflection transforms matter into antimatter: the electronic ring current becomes a positronic ring current, the radioactive cobalt is replaced by radioactive anticobalt.

If it is true that a symmetry plane always remains a symmetry plane, the initial state of the Co experiment could not have contained a symmetry plane. This would not be the case if the magnetic vector were polar—in which case the electric vector would be axial. The charge density, the divergence of the electric vector, would then become a pseudoscalar rather than a simple scalar as in current theory. The mirror image of a negative charge would be positive, the mirror image of an electron a positron, and conversely. The mirror image of matter would be antimatter. The Co experiment, viewed through a mirror, would not present a picture contrary to established fact: it would present an experiment carried out with antimatter. The right side of Fig. 3 shows the mirror image of the left side. Thus, the principle of sufficient cause, and the validity of symmetry planes, need not be abandoned if one is willing to admit that the mirror image of matter is antimatter.

The possibility just envisaged would be technically described as the elimination of the operations of reflection and charge conjugation, as presently defined, as true symmetry operations. Their product would still be assumed to be a symmetry operation and proposed to be named, simply, reflection. A few further technical remarks are contained in Appendix II. The proposition just made has two aspects: a very appealing one, and a very alarming one.

Let us look first at the appealing aspect. Dirac has said that the number of elementary particles shows an alarming tendency of increasing. One is tempted to add to this that the number of invariance properties

---

[5] Lee, Yang, and Oehme, Phys. Rev. **106**, 340 (1957).

[5a] The interpretation referred to has been proposed independently by numerous authors, including A. Salam, Nuovo cimento **5**, 229 (1957); L. Landau, Nuclear Phys. **3**, 127 (1957); H. D. Smyth and L. Biedenharn (personal communication). Dr. S. Deser has pointed out that the "perturbing possibility" was raised already by Wick, Wightman, and Wigner [Phys. Rev. **88**, 101 (1952)] but was held "remote at that time." Naturally, the apparent unanimity of opinion does not prove its correctness.

also showed a similar tendency. It is not equally alarming because, while the increase in the number of elementary particles complicates our picture of nature, that of the symmetry properties on the whole simplifies it. Nevertheless the clear correspondence between the invariance properties of the laws of nature, and the symmetry properties of space-time, was most clearly breached by the operation of charge conjugation. This postulated that the laws of nature remain the same if all positive charges are replaced by negative charges and vice versa, or more generally, if all particles are replaced by antiparticles. Reasonable as this postulate appears to us, it corresponds to no symmetry of the space-time continuum. If the preceding interpretation of the Co experiments should be sustained, the correspondence between the natural symmetry elements of space-time, and the invariance properties of the laws of nature, would be restored. It is true that the role of the planes of reflection would not be that to which we are accustomed—the mirror image of an electron would become a positron—but the mirror image of a sequence of events would still be a possible sequence of events. This possible sequence of events would be more difficult to realize in the actual physical world than what we had thought, but it would still be possible.

The restoration of the correspondence between the natural symmetry properties of space-time on one hand, and the laws of nature on the other hand, is the appealing feature of the proposition. It has, actually, two alarming features. The first of these is that a symmetry operation is, physically, so complicated. If it should turn out that the operation of time inversion, as we now conceive it, is not a valid symmetry operation (e.g., if one of the experiments proposed by Treiman and Wyld gave a positive result) we could still maintain the validity of this symmetry operation by reinterpreting it. We could postulate, for instance, that time inversion transforms matter into *meta*-matter which will be discovered later when higher energy accelerators will become available. Thus, maintaining the validity of symmetry planes forces us to a more artificial view of the concept of symmetry and of the invariance of the laws of physics.

The other alarming feature of our new knowledge is that we have been misled for such a long time to believe in more symmetry elements than actually exist. There was ample reason for this and there was ample experimental evidence to believe that the mirror image of a possible event is again a possible event with electrons being the mirror images of electrons and not of positrons. Let us recall in this connection first how the concept of parity, resulting from the beautiful though almost forgotten experiments of Laporte,[6] appeared to be a perfectly valid concept in spectroscopy and in nuclear physics. This concept could be explained very naturally as a result of the reflection symmetry of space-time, the mirror image of electrons being electrons and not positrons. We are now forced to believe that this symmetry is only approximate and the concept of parity, as used in spectroscopy and nuclear physics, is also only approximate. Even more fundamentally, there is a vast body of experimental information in the chemistry of optically active substances which are mirror images of each other and which have optical activities of opposite direction but exactly equal strength. There is the fact that molecules which have symmetry planes are optically inactive; there is the fact of symmetry planes in crystals.[7] All these facts relate properties of right-handed matter to left-handed *matter*, not of right-handed matter to left-handed *antimatter*. The new experiments leave no doubt that the symmetry plane in this sense is not valid for all phenomena, in particular not valid for β decay, that if the concept of symmetry plane is at all valid for all phenomena, it can be valid only in the sense of converting matter into antimatter.

Furthermore, the old-fashioned type of symmetry plane is not the only symmetry concept that is only approximately valid. Charge conjugation was mentioned before, and we are remainded also of isotopic spin, of the exchange character, that is multiplet system, for electrons and also of nuclei which latter holds so accurately that, in practice, parahydrogen molecules can be converted into orthohydrogen molecules only by first destroying them.[8] This approximate validity of laws of symmetry is, therefore, a very general phenomenon—it may be *the* general phenomenon. We are reminded of Mach's axiom that the laws of nature depend on the physical content of the universe, and the physical content of the universe certainly shows no symmetry. This suggests—and this may also be the spirit of the ideas of Yang and Lee—that all symmetry properties are only approximate. The weakest interaction, the gravitational force, is the basis of the distinction between inertial and accelerated coordinate systems, the second weakest known interaction, that leading to β decay, leads to the distinction between matter and antimatter. Let me conclude this subject by expressing the conviction that the discoveries of Wu, Ambler, Hayward, Hoppes, and Hudson,[9] and of Garwin, Lederman, and Weinreich[10] will not remain isolated discoveries. More likely, they herald a revision of our concept of invariance and possibly

[6] O. Laporte, Z. Physik 23, 135 (1924). For the interpretation of Laporte's rule in terms of the quantum-mechanical operation of inversion, see the writer's *Gruppentheorie und ihre Anwendungen auf die Quantenmechanik der Atmospektren* (Friedrich Vieweg und Sohn, Braunschweig, 1931), Chap. XVIII.

[7] For the role of the space and time inversion operators in classical theory, see H. Zocher and C. Török, Proc. Natl. Acad. Sci. U.S. 39, 681 (1953) and literature quoted there.

[8] See A. Farkas, *Orthohydrogen, Parahydrogen and Heavy Hydrogen* (Cambridge University Press, New York, 1935).

[9] Wu, Ambler, Hayward, Hoppes, and Hudson, Phys. Rev. **105**, 1413(L) (1957).

[10] Garwin, Lederman, and Weinreich, Phys. Rev. **105**, 1415(L) (1957); also, J. L. Friedman and V. L. Telegdi, *ibid.* **105**, 1681(L) (1957).

of other concepts which are even more taken for granted.

QUANTUM LIMITATIONS OF THE CONCEPTS OF GENERAL RELATIVITY

The last remarks naturally bring us to a discussion of the general theory of relativity. The main premise of this theory is that coordinates are only labels to specify space-time points. Their values have no particular significance unless the coordinate system is somehow anchored to events in space-time.

Let us look at the question of how the equations of the general theory of relativity could be verified. The purpose of these equations, as of all equations of physics, is to calculate, from the knowledge of the present, the state of affairs that will prevail in the future. The quantities describing the present state are called initial conditions; the ways these quantities change are called the equations of motion. In relativity theory, the state is described by the metric which consists of a network of points in space-time, that is a network of events, and the distances between these events. If we wish to translate these general statements into something concrete, we must decide what events are, and how we measure distances between *events*. The metric in the general theory of relativity is a metric in space-time, its elements are distances between space-time points, not between points in ordinary space.

The events of the general theory of relativity are coincidences, that is, collisions between particles. The founder of the theory, when he created this concept, had evidently macroscopic bodies in mind. Coincidences, that is, collisions between such bodies, are immediately observable. This is not the case for elementary particles; a collision between these is something much more evanescent. In fact, the point of a collision between two elementary particles can be closely localized in space-time only in case of high-energy collisions. (See

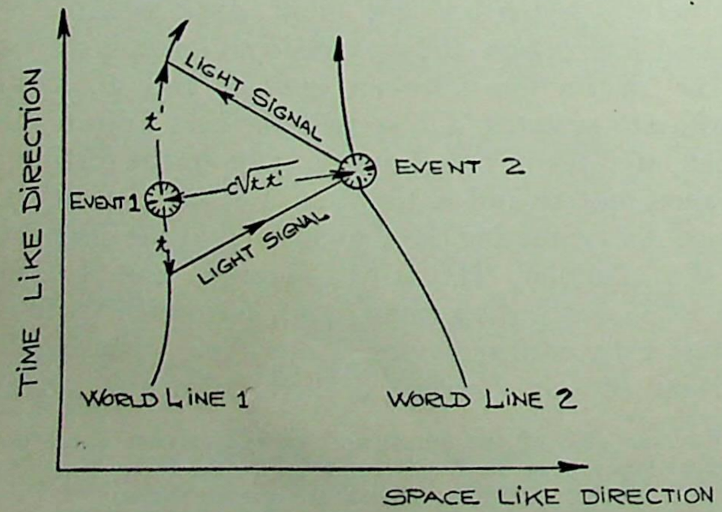


Fig. 4. Measurement of space-like distances by means of a clock. It is assumed that the metric tensor is essentially constant within the space-time region contained in the figure. The space-like distance between events 1 and 2 is measured by means of the light signals which pass through event 2 and a geodesic which goes through event 1. Explanation in Appendix IV.

Appendix III.) This shows that the establishment of a close network of points in space-time requires a reasonable energy density, a dense forest of world lines wherever the network is to be established. However, it is not necessary to discuss this in detail because the measurement of the distances between the points of the network gives more stringent requirements than the establishment of the network.

It is often said that the distances between events must be measured by yardsticks and rods. We found that measurements with a yardstick are rather difficult to describe and that their use would involve a great deal of unnecessary complications. The yardstick gives the distance between events correctly only if its marks coincide with the two events simultaneously from the point of view of the rest-system of the yardstick. Furthermore, it is hard to image yardsticks as anything but macroscopic objects. It is desirable, therefore, to reduce all measurements in space-time to measurements by clocks. Naturally, one can measure by clocks directly only the distances of points which are in time-like relation to each other. The distances of events which are in space-like relation, and which would be measured more naturally by yardsticks, will have to be measured, therefore, indirectly.

It appears, thus, that the simplest framework in space-time, and the one which is most nearly microscopic, is a set of clocks, which are only slowly moving with respect to each other, that is, with world lines which are approximately parallel. These clocks tick off periods and these ticks form the network of events which we wanted to establish. This, at the same time, establishes the distance of those adjacent points which are on the same world line.

Figure 4 shows two world lines and also shows an event, that is, a tick of the clock, on each. The figure shows an artifice which enables one to measure the distance of space-like events: a light signal is sent out from the first clock which strikes the second clock at event 2. This clock, in turn, sends out a light signal which strikes the first clock at time t' after the event 1. If the first light signal had to be sent out at time t before the first event, the calculation given in Appendix IV shows that the space-like distance of events 1 and 2 is the geometric average of the two measured time-like distances t and t'. This is then a way to measure distances between space-like events by clocks instead of yardsticks.

It is interesting to consider the quantum limitations on the accuracy of the conversion of time-like measurements into space-like measurements, which is illustrated in Fig. 4. Naturally, the times t and t' will be well defined only if the light signal is a short pulse. This implies that it is composed of many frequencies and, hence, that its energy spectrum has a corresponding width. As a result, it will give an indeterminate recoil to the second clock, thus further increasing the uncertainty of its momentum. All this is closely related

to Heisenberg's uncertainty principle. A more detailed calculation[1] shows that the added uncertainty is of the same order of magnitude as the uncertainty inherent in the nature of the best clock that we could think of, so that the conversion of time-like measurements into space-like measurements is essentially free.

We finally come to the discussion of one of the principal problems—the limitations on the accuracy of the clock. It led us to the conclusion that the inherent limitations on the accuracy of a clock of given weight and size, which should run for a period of a certain length, are quite severe. In fact, the result in summary is that a clock is an essentially nonmicroscopic object. In particular, what we vaguely call an atomic clock, a single atom which ticks off its periods, is surely an idealization which is in conflict with fundamental concepts of measurability. This part of our conclusions can be considered to be well established. On the other hand, the actual formula which will be given for the limitation of the accuracy of time measurement, a sort of uncertainty principle, should be considered as the best present estimate.

Let us state the requirements as follows. The watch shall run T seconds, shall measure time with an accuracy of $T/n=t$, its linear extension shall not exceed l, its mass shall be below m. Since the pointer of the watch must be able to assume n different positions, the system will have to run, in the course of the time T, over at least n orthogonal states. Its state must, therefore, be the superposition of at least n stationary states. It is clear, furthermore, that unless its total energy is at least h/t, it cannot measure a time interval which is smaller than t. This is equivalent to the usual uncertainty principle. These two requirements follow directly from the basic principles of quantum theory; they are also the requirements which could well have been anticipated. A clock which conforms with these postulates is, for instance, an oscillator, with a period which is equal to the running time of the clock, if it is with equal probabilty in any of the first n quantum states. Its energy is about n times the energy of the first excited state. This corresponds to the uncertainty principle with the accuracy t as time uncertainty. Broadly speaking, the clock is a very soft oscillator, the oscillating particle moving very slowly and with a rather large amplitude. The pointer of the clock is the position of the oscillating particle.

The clock of the preceding paragraph is still very light. Let us consider, however, the requirement that the linear dimensions of the clock be limited. Since there is little point in dealing with the question in great generality, it may as well be assumed here that the linear dimension shall correspond to the accuracy in time. The requirement $l=ct$ increases the mass of the clock by n^3 which may be a very large factor indeed:

$$m>n^3\hbar t/l^2=n^3\hbar/c^2t. \qquad (2)$$

For example, a clock, with a running time of a day and an accuracy of 10^{-8} second, must weigh almost a gram—for reasons stemming solely from uncertainty principles and similar considerations.

So far, we have paid attention only to the physical dimension of the clock and the requirement that it be able to distinguish between events which are only a distance t apart on the time scale. In order to make it usable as part of the framework which was described before, it is necessary to *read* the clock and to start it. As part of the framework to map out the metric of space-time, it must either register the readings at which it receives impulses, or transmit these readings to a part of space outside the region to be mapped out. This point was already noted by Schrödinger.[11] However, we found it reassuring that, in the most interesting case in which $l=ct$, that is, if space and time inaccuracies are about equal, the reading requirement introduces only an insignificant numerical factor but does not change the form of the expression for the minimum mass of the clock.

The arrangement to map the metric might consist, therefore, of a lattice of clocks, all more or less at rest with respect to each other. All these clocks can emit light signals and receive them. They can also transmit their reading at the time of the receipt of the light signal to the outside. The clocks may resemble oscillators, well in the nonrelativistic region. In fact, the velocity of the oscillating particle is about n times smaller than the velocity of light where n is the ratio of the *error* in the time measurement, to the *duration* of the whole interval to be measured. This last quantity is the spacing of the events on the time axis, it is also the distance of the clocks from each other, divided by the light velocity. The world lines of the clocks from the dense forest which was mentioned before. Its branches suffuse the region of space-time in which the metric is to be mapped out.

We are not absolutely convinced that our clocks are the best possible. Our principal concern is that we have considered only one space-like dimension. One consequence of this was that the oscillator had to be a one-dimensional oscillator. It is possible that the size limitation does not increase the necessary mass of the clock to the same extent if use is made of all three spatial dimensions.

The curvature tensor can be obtained from the metric in the conventional way, if the metric is measured with sufficient accuracy. It may be of interest, nevertheless, the describe a more direct method for measuring the curvature of space. It involves an arrangement, illustrated in Fig. 5, which is similar to that used for obtaining the metric. There is a clock, and a mirror, at such a distance from each other that the curvature of space can be assumed to be constant in the interven-

[11] E. Schrödinger, Ber. Preuss. Akad. Wiss. phys.-math. Kl. 1931, 238.

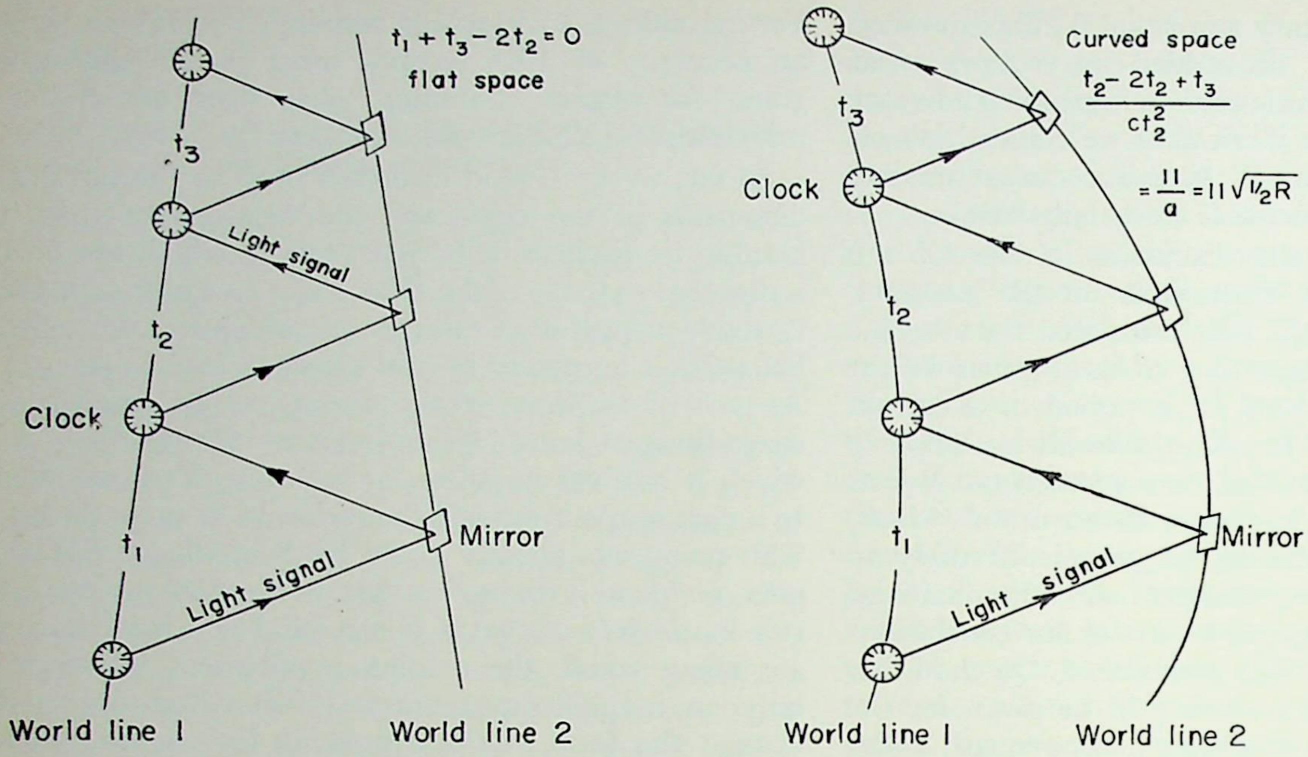


FIG. 5. Direct measurement of the curvature by means of a clock and mirror. Only one space-like dimension is considered and the curvature assumed to be constant within the space-time region contained in the figure. The explanation is given in Appendix V.

ing region. The two clocks need not be at rest with respect to each other, in fact, such a requirement would involve additional measurements to verify it. If the space is flat, the world lines of the clocks can be drawn straight. In order to measure the curvature, a light signal is emitted by the clock, and this is reflected by the mirror. The time of return is read on the clock—it is t_1—and the light signal returned to the mirror. The time which the light signal takes on its second trip to return to the clock is denoted by t_2. The process is repeated a third time, the duration of the last roundtrip denoted by t_3. As shown in Appendix V, the radius of curvature, a, and the relevant component R_{0101} of the Riemann tensor are given by

$$\frac{t_1 - 2t_2 + t_3}{t_2^2} = \frac{11}{a} = 11(\tfrac{1}{2}R_{0101})^{\frac{1}{2}}. \tag{2}$$

If classical theory would be valid also in the microscopic domain, there would be no limit on the accuracy of the measurement indicated in Fig. 5. If $\hbar$ is infinitely small, the time intervals t_1, t_2, t_3 can all be measured with arbitrary accuracy with an infinitely light clock. Similarly, the light signals between clock and mirror, however short, need carry only an infinitesimal amount of momentum and thus deflect clock and mirror arbitrarily little from their geodesic paths. The quantum phenomena considered before force us, however, to use a clock with a minimum mass if the measurement of the time intervals is to have a given accuracy. In the present case, this accuracy must be relatively high unless the time intervals t_1, t_2, t_3 are of the same order of magnitude as the curvature of space. Similarly, the deflection of clock and mirror from their geodesic paths must be very small if the result of the measurement is to be meaningful. This gives an effective limit for the accuracy with which the curvature can be measured. The result is, as could be anticipated, that the curvature at a *point* in space-time cannot be measured at all; only the average curvature over a finite region of space-time can be obtained. The error of the measurement[1] is inversely proportional to the two-thirds power of the area available in space-time, that is, the area around which a vector is carried, always parallel to itself, in the customary definition of the curvature. The error is also proportional to the cube root of the Compton wavelength of the clock. Our principal hesitation in considering this result as definitive is again its being based on the consideration of only one space-like dimension. The possibilities of measuring devices, as well as the problems, may be substantially different in three-dimensional space.

Whether or not this is the case, the essentially nonmicroscopic nature of the general relativistic concepts seems to us inescapable. If we look at this first from a practical point of view, the situation is rather reassuring. We can note first, that the measurement of electric and magnetic fields, as discussed by Bohr and Rosenfeld,[12] also requires macroscopic, in fact *very* macroscopic, equipment and that this does not render the electromagnetic field concepts useless for the purposes of quantum electrodynamics. It is true that the measurement of space-time curvature requires a finite region of space and there is a minimum for the mass, and even the mass uncertainty, of the measuring equipment. However, numerically, the situation is by no means alarming. Even in interstellar space, it should be possible to measure the curvature

[12] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab Mat.-fys. Medd. 12, No. 8 (1933). See also further literature quoted in L. Rosenfeld's article in *Niels Bohr and the Development of Physics* (Pergamon Press, London, 1955).

in a volume of a light second or so. Furthermore, the mass of the clocks which one will wish to employ for such a measurement is of the order of several micrograms so that the finite mass of elementary particles does not cause any difficulty. The clocks will contain many particles and there is no need, and there is not even an incentive, to employ clocks which are lighter than the elementary particles. This is hardly surprising since the mass which can be derived from the gravitational constant, light velocity, and Planck's constant, is about 20 micrograms.

It is well to repeat, however, that the situation is less satisfactory from a more fundamental point of view. It remains true that we consider, in ordinary quantum theory, position operators as observables without specifying what the coordinates mean. The concepts of quantum field theories are even more weird from the point of view of the basic observation that only coincidences are meaningful. This again is hardly surprising because even a 20-microgram clock is too large for the measurement of atomic times or distances. If we analyze the way in which we "get away" with the use of an absolute space concept, we simply find that we do not. In our experiments we surround the microscopic objects with a very macroscopic framework and observe *coincidences* between the particles emanating from the microscopic system, and parts of the framework. This gives the collision matrix, which is observable, and observable in terms of macroscopic coincidences. However, the so-called observables of the microscopic system are not only not observed, they do not even appear to be meaningful. There is, therefore, a boundary in our experiments between the region in which we use the quantum concepts without worrying about their meaning in face of the fundamental observation of the general theory of relativity, and the surrounding region in which we use concepts which are meaningful also in the face of the basic observation of the general theory of relativity but which cannot be described by means of quantum theory. This appears most unsatisfactory from a strictly logical standpoint.

APPENDIX I

It will be necessary, in this appendix, to compare various states of the same physical system. These states will be generated by looking at the same state—the standard state—from various coordinate systems. Hence every Lorentz frame of reference will define a state of the system—the state as which the standard state appears from the point of view of this coordinate system. In order to define the standard state, we choose an arbitrary but fixed Lorentz frame of reference and stipulate that, in this frame of reference, the particle in the standard state be at rest and its spin (if any) have the direction of the z axis. Thus, if we wish to have a particle moving with a velocity v in the z direction and with a spin also directed along this axis, we look at the particle in the standard state from a coordinate system moving with the velocity v in the $-z$ direction. If we wish to have a particle at rest but with its spin in the yz plane, including an angle α with the z axis, we look at the standard state from a coordinate system the y and z axes of which include an angle α with the y and z axes of the coordinate system in which the standard state was defined. In order to obtain a state in which both velocity and spin have the aforementioned direction (i.e., a direction in the yz plane, including the angles α and $\frac{1}{2}\pi-\alpha$ with the y and z axes), we look at the standard state from the point of view of a coordinate system in which the spin of the standard state is described as this direction and which is moving in the opposite direction.

Two states of the system will be identical only if the Lorentz frames of reference which define them are identical. Under this definition, the relations which will be obtained will be valid independently of the properties of the particle, such as spin or mass (as long as the mass in nonzero so that the standard state exists). Two states will be approximately the same if the two Lorentz frames of reference which define them can be obtained from each other by a very small Lorentz transformation, that is, one which is near the identity. Naturally, all states of a particle which can be compared in this way are related to each other inasmuch as they represent the same standard state viewed from various coordinate systems. However, we shall have to compare only these states.

Let us denote by $A(0,\varphi)$ the matrix of the transformation in which the transformed coordinate system moves with the velocity $-v$ in the z direction where $v=c\tanh\varphi$

$$A(0,\varphi)=\left\|\begin{matrix}1 & 0 & 0\\ 0 & \cosh\varphi & \sinh\varphi\\ 0 & \sinh\varphi & \cosh\varphi\end{matrix}\right\|. \tag{1.1}$$

Since the x axis will play no role in the following consideration, it is suppressed in (1.1) and the three rows and the three columns of this matrix refer to the y', z', ct' and to the y, z, ct axes, respectively. The matrix (1.1) characterizes the state in which the particle moves with a velocity v in the direction of the z axis and its spin is parallel to this axis.

Let us further denote the matrix of the rotation by an angle φ in the yz plane by

$$R(\vartheta)=\left\|\begin{matrix}\cos\vartheta & \sin\vartheta & 0\\ -\sin\vartheta & \cos\vartheta & 0\\ 0 & 0 & 1\end{matrix}\right\|. \tag{1.2}$$

We refer to the direction in the yz plane which lies between the y and z axes and includes an angle ϑ with the z axis as the direction ϑ. The coordinate system which moves with the velocity $-v$ in the ϑ direction is obtained by the transformation

$$A(\vartheta,\varphi)=R(\vartheta)A(0,\varphi)R(-\vartheta). \tag{1.3}$$

In order to obtain a particle which moves in the direction ϑ and is polarized in this direction, we first rotate the coordinate system counterclockwise by ϑ (to have the particle polarized in the proper direction) and impart it then a velocity $-v$ in the ϑ direction. Hence, it is the transformation

$$T(\vartheta,\varphi)=A(\vartheta,\varphi)R(\vartheta)$$
$$=\left\|\begin{matrix}\cos\vartheta & \sin\vartheta\cosh\varphi & \sin\vartheta\sinh\varphi\\ -\sin\vartheta & \cos\vartheta\cosh\varphi & \cos\vartheta\sinh\varphi\\ 0 & \sinh\varphi & \cosh\varphi\end{matrix}\right\| \quad (1.4)$$

which characterizes the aforementioned state of the particle. It follows from (1.3) that

$$T(\vartheta,\varphi)=R(\vartheta)A(0,\varphi)=R(\vartheta)T(0,\varphi) \quad (1.5)$$

so that the same state can be obtained also by viewing the state characterized by (1.1) from a coordinate system that is rotated by ϑ. It follows that the statement "velocity and spin are parallel" is invariant under rotations. This had to be expected.

If the state generated by $A(0,\varphi)=T(0,\varphi)$ is viewed from a coordinate system which is moving with the velocity u in the direction of the z axis, the particle will still appear to move in the z direction and its spin will remain parallel to its direction of motion, unless $u>v$ in which case the two directions will become antiparallel, or unless $u=v$ in which case the statement becomes meaningless, the particle appearing to be at rest. Similarly, the other states in which spin and velocity are parallel, i.e., the states generated by the transformations $T(\vartheta,\varphi)$, remain such states if viewed from a coordinate system moving in the direction of the particle's velocity, as long as the coordinate system is not moving faster than the particle. This also had to be expected. However, if the state generated by $T(0,\varphi)$ is viewed from a coordinate system moving with velocity $v'=c\tanh\varphi'$ in the $-y$ direction, spin and velocity will *not* appear parallel any more, *provided the velocity v of the particle is not close to light velocity.* This last proviso is the essential one; it means that the high velocity states of a particle for which spin and velocity are parallel (i.e., the states generated by (1.4) with a large φ) are states of this same nature if viewed from a coordinate system which is not moving too fast in the direction of motion of the particle itself. In the limiting case of the particle moving with light velocity, the aforementioned states become invariant under *all* Lorentz transformations.

Let us first convince ourselves that if the state (1.1) is viewed from a coordinate system moving in the $-y$ direction, its spin and velocity no longer appear parallel. The state in question is generated from the normal state by the transformation

$$A(\tfrac{1}{2}\pi,\varphi')A(0,\varphi)$$
$$=\left\|\begin{matrix}\cosh\varphi' & \sinh\varphi\sinh\varphi' & \cosh\varphi\sinh\varphi'\\ 0 & \cosh\varphi & \sinh\varphi\\ \sinh\varphi' & \sinh\varphi\cosh\varphi' & \cosh\varphi\cosh\varphi'\end{matrix}\right\|. \quad (1.6)$$

This transformation does not have the form (1.4). In order to bring it into that form, it has to be multiplied on the right by $R(\epsilon)$, i.e., one has to rotate the spin ahead of time. The angle ϵ is given by the equation

$$\tan\epsilon=\frac{\tanh\varphi'}{\sinh\varphi}=\frac{v'}{v}(1-v^2/c^2)^{\frac{1}{2}} \quad (1.7)$$

and is called the angle between spin and velocity. For $v\ll c$, it becomes equal to the angle which the ordinary resultant of two perpendicular velocities, v and v', includes with the first of these. However, ϵ becomes very small if v is close to c; in this case it is hardly necessary to rotate the spin away from the z axis before giving it a velocity in the z direction. These statements express the identity

$$A(\tfrac{1}{2}\pi,\varphi')A(0,\varphi)R(\epsilon)=T(\vartheta,\varphi'') \quad (1.8)$$

which can be verified by direct calculation. The right side represents a particle with parallel spin and velocity, the magnitude and direction of the latter being given by the well-known equations

$$v''=c\tanh\varphi''=(v^2+v'^2-v^2v'^2/c^2)^{\frac{1}{2}} \quad (1.8a)$$

and

$$\tan\vartheta=\frac{\sinh\varphi'}{\tanh\varphi}=\frac{v'}{v(1-v'^2/c^2)^{\frac{1}{2}}}. \quad (1.8b)$$

Equation (1) given in the text follows from (1.7) and (1.8b) for $v\sim c$.

The fact that the states $T(\vartheta,\varphi)\psi_0$ (where ψ_0 is the standard state and $\varphi\gg1$) are approximately invariant under all Lorentz transformations is expressed mathematically by the equations,

$$R(\vartheta)\cdot T(0,\varphi)\psi_0=T(\vartheta,\varphi)\psi_0, \quad (1.5a)$$

$$A(0,\varphi')\cdot T(0,\varphi)\psi_0=T(0,\varphi'+\varphi)\psi_0, \quad (1.9a)$$

and

$$A(\tfrac{1}{2}\pi,\varphi')\cdot T(0,\varphi)\psi_0\rightarrow T(\vartheta,\varphi'')\psi_0, \quad (1.9b)$$

which give the wave function of the state $T(0,\varphi)\psi_0$, as viewed from other Lorentz frames of reference. Naturally, similar equations apply to all $T(\alpha,\varphi)\psi_0$. In particular, (1.5a) shows that the states in question are invariant under rotations of the coordinate system, (1.9a) that they are invariant with respect to Lorentz transformations with a velocity not too high *in* the direction of motion (so that $\varphi'+\varphi\gg0$, i.e., φ' not too large a negative number). Finally, in order to prove (1.9b), we calculate the transition probability between the states $A(\frac{1}{2}\pi,\varphi')\cdot T(0,\varphi)\psi_0$ and $T(\vartheta,\varphi'')\psi_0$ where ϑ and φ'' are given by (1.8a) and (1.8b). For this, (1.8) gives

$$(A(\tfrac{1}{2}\pi,\varphi')\cdot T(0,\varphi)\psi_0,\ T(\vartheta,\varphi'')\psi_0)$$
$$=(T(\vartheta,\varphi'')R(\epsilon)^{-1}\psi_0,\ T(\vartheta,\varphi'')\psi_0)$$
$$=(R(\epsilon)^{-1}\psi_0,\psi_0)\rightarrow(\psi_0,\psi_0).$$

The second line follows because $T(\vartheta,\varphi'')$ represents a coordinate transformation and is, therefore, unitary. The last member follows because $\epsilon \rightarrow 0$ as $\varphi \rightarrow \infty$ as can be seen from (1.7) and $R(0)=1$.

The preceding consideration is not fundamentally new. It is an elaboration of the facts (a) that the subgroup of the Lorentz group which leaves a null-vector invariant is different from the subgroup which leaves a time like vector invariant[4] and (b) that the representations of the latter subgroup decompose into one dimensional representations if this subgroup is "contracted" into the subgroup which leaves a null-vector invariant.[13]

APPENDIX II

Before the hypothesis of Lee and Yang[14] was put forward, it was commonly assumed that there are, in addition to the symmetry operations of the proper Poincaré group, three further independent symmetry operations. The proper Poincaré group consists of all Lorentz transformations which can be continuously obtained from unity and all translations in space-like and time-like directions, as well as the products of all these transformations. It is a continuous group; the Lorentz transformations contained in it do not change the direction of the time axis and their determinant is 1. The three independent further operations which were considered to be rigorously valid, were

Space inversion I, that is, the transformation $x, y, z \rightarrow -x, -y, -z$, without changing particles into antiparticles.

Time inversion T, more appropriately described by Lüders[15] as *Umkehr der Bewegungsrichtung*, which replaces every velocity by the opposite velocity so that the position of the particles at $+t$ becomes the same as it was, without time inversion, at $-t$. The time inversion T (also called time inversion of the first kind by Lüders[16]) does not convert particles into antiparticles either.

Charge conjugation C, that is, the replacement of positive charges by negative charges and more generally of particles by antiparticles, without changing either the position or the velocity of these particles.[17] The quantum-mechanical expressions for the symmetry operations I and C are unitary, that for T is antiunitary.

The three operations I, T, C, together with their products TC (Lüders' time inversion of the second kind), IC, IT, ITC and the unit operation form a group and the products of the elements of this group with those of the proper Poincaré group were considered to be the symmetry operations of all laws of physics. The suggestion given in the text amounts to eliminating the operations I and C separately while continuing to postulate their product IC as symmetry operation. The discrete symmetry group then reduces to the unit operation plus

$$IC,\ T,\ \text{and}\ ICT, \tag{2.1}$$

and the total symmetry group of the laws of physics becomes the proper Poincaré group plus its products with the elements (2.1). This group is isomorphic (essentially identical) with the unrestricted Poincaré group, i.e., the product of *all* Lorentz transformations with all the displacements in space and time. The quantum mechanical expressions for the operations of the proper Lorentz group and its product with IC are unitary, those for T and ICT (as well as for their products with the elements of the proper Poincaré group) antiunitary. Lüders[16] has pointed out that, under certain very natural conditions, ICT belongs to the symmetry group of every *local* field theory.

APPENDIX III

Let us consider, first, the collision of two particles of equal mass m in the coordinate system in which the average of the sum of their momenta is zero. Let us assume that, at a given time, the wave function of both particles is confined to a distance l in the direction of their average velocity with respect to each other. If we consider only this space-like direction, and the time axis, the area in space-time in which the two wave functions will substantially overlap is [see Fig. 6(a)]

$$a = l^2/2v_{\min}, \tag{3.1}$$

where $v_{\min}$ is the lowest velocity which occurs with substantial probability in the wave packets of the colliding particles. Denoting the average momentum by $\bar{p}$ (this has the same value for both particles) the half-width of the momentum distribution by δ, then $v_{\min} = (\bar{p}-\delta)(m^2+(\bar{p}-\delta)^2/c^2)^{-\frac{1}{2}}$. Since l cannot be below $\hbar/\delta$, the area (3.1) is at least

$$\frac{\hbar^2}{2\delta^2}\frac{(m^2+(\bar{p}-\delta)^2/c^2)^{\frac{1}{2}}}{\bar{p}-\delta}. \tag{3.1a}$$

(Note that the area becomes infinite if $\delta \geq \bar{p}$.) The

---

[13] E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953).

[14] T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956). See also E. M. Purcell and N. F. Ramsey, Phys. Rev. **78**, 807 (1950).

[15] G. Lüders, Z. Physik **133**, 325 (1952).

[16] G. Lüders, Kgl. Danske Videnskab. Selskab Mat.-fys. Medd. **28**, No. 5 (1954).

[17] All three symmetry operations were first discussed in detail by J. Schwinger, Phys. Rev. **74**, 1439 (1948). See also H. A. Kramers, Proc. Acad. Sci. Amsterdam **40**, 814 (1937) and W. Pauli's article in *Niels Bohr and the Development of Physics* (Pergamon Press, London, 1955). The significance of the first two symmetry operations (and their connection with the concepts of parity and the Kramers degeneracy respectively), were first pointed out by the present writer, Z. Physik **43**, 624 (1927) and Nachr. Akad. Wiss. Göttingen, Math.-physik. **1932**, 546. See also T. D. Newton and E. P. Wigner, Revs. Modern Phys. **21**, 400 (1949); S. Watanabe, Revs. Modern Phys. **27**, 26 (1945). The concept of charge conjugation is based on the observation of W. Furry, Phys. Rev. **51**, 125 (1937).

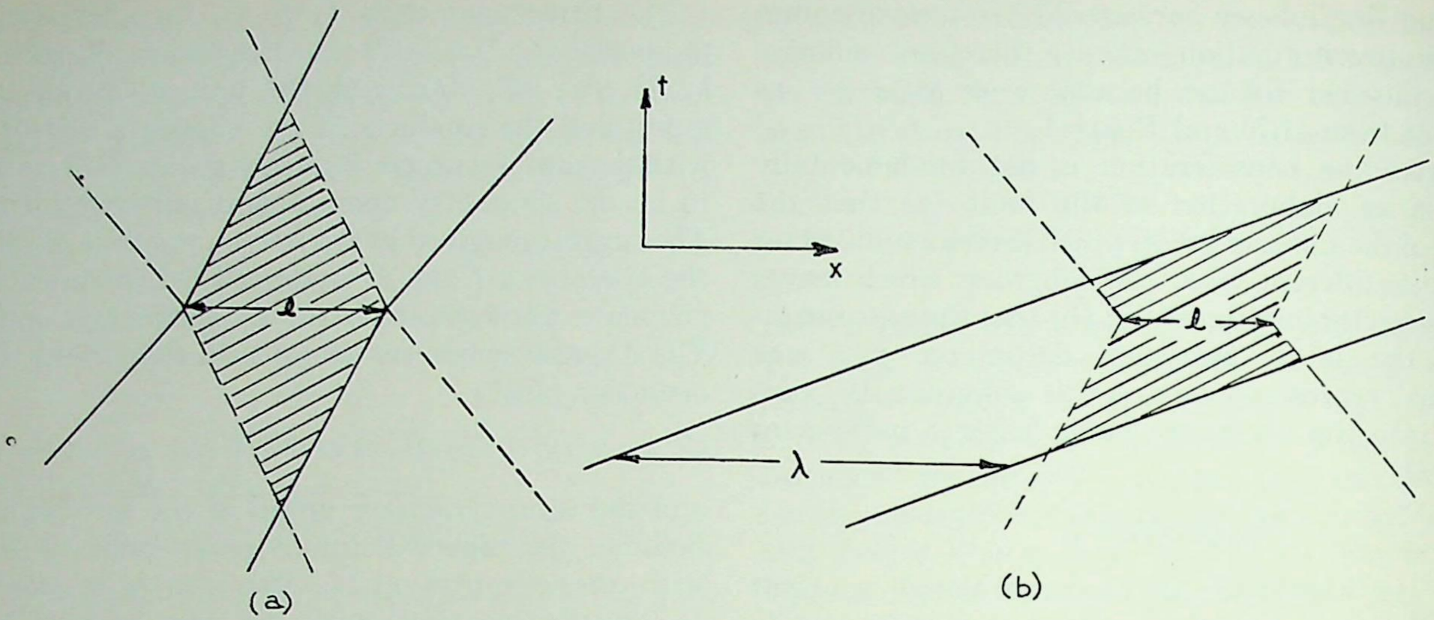


FIG. 6. (a) Localization of a collision of two particles of equal mass. The full lines indicate the effective boundaries of the wave packet of the particle traveling to the right, the broken lines the effective boundaries of the wave packet of the particle traveling to the left. The collision can take place in the shaded area of space-time. (b) Localization of a collision between a particle with finite mass and a particle with zero rest-mass. The full lines, at a distance λ apart in the x direction, indicate the boundary of the particle with zero rest-mass, the broken lines apply to the wave packet of the particle with nonzero rest-mass. The collision can take place in the shaded area.

minimum of (3.1a) is, apart from a numerical factor

$$a_{\min}\approx\frac{\hbar^2}{\bar{p}^3}(m^2+\bar{p}^2/c^2)^{\frac{1}{2}}\approx\frac{\hbar^2 c}{E^{\frac{3}{2}}(E+mc^2)^{\frac{1}{2}}}, \tag{3.2}$$

where E is the kinetic energy (total energy minus rest-energy) of the particles.

The kinetic energy E permits the contraction of the wave functions of the colliding particles also in directions perpendicular to the average relative velocity, to an area $\hbar^2c^2/E(E+2mc^2)$. Hence, again apart from a numerical factor, the volume to which the collision can be confined in four dimensional space-time becomes

$$V_{\min}=\frac{\hbar^4c^3}{E^{\frac{5}{2}}(E+mc^2)^{\frac{3}{2}}}. \tag{3.3}$$

E is the average kinetic energy of the particles in the coordinate system in which their center of mass is, on the average, at rest. Equation (3.3) is valid apart from a numerical constant of unit order of magnitude but this constant depends on E/mc^2.

Let us consider now the opposite limiting case, the collision of a particle with finite rest-mass m with a particle with zero rest-mass. The collision is viewed again in the coordinate system in which the average linear momentum is zero. In this case, one will wish to confine the wave function of the particle with finite rest-mass to a narrower region l than that of the particle with zero rest-mass. If the latter is confined to a region of thickness λ, [see Fig. 6(b)], its momentum and energy uncertainties will be at least $\hbar/\lambda$ and $\hbar c/\lambda$ and these expressions will also give, apart from a numerical factor, the average values of these quantities. Hence $\bar{p}\approx\hbar/\lambda$. The kinetic energy of the particle with finite restmass will be of the order of magnitude

$$\tfrac{1}{2}(m^2c^4+(\bar{p}+\hbar/l)^2c^2)^{\frac{1}{2}}+\tfrac{1}{2}(m^2c^4+(\bar{p}-\hbar/l)^2c^2)^{\frac{1}{2}}-mc^2, \tag{3.4}$$

since $\hbar/l$ is the momentum uncertainty. Since $l\leq\lambda$, one can neglect $\bar{p}$ in (3.4) if one is interested only in the order of magnitude. This gives for the total kinetic energy,

$$E\approx\hbar c/\lambda+(m^2c^4+\hbar^2c^2/l^2)^{\frac{1}{2}}-mc^2, \tag{3.5}$$

while the area in Fig. 6(b) is of the order of magnitude

$$a=(\lambda/c)(l+\Delta v\lambda/c), \tag{3.6}$$

where Δv is the uncertainty in the velocity of the second particle

$$\Delta v=\frac{\bar{p}+\hbar/l}{(m^2+(\bar{p}+\hbar/l)^2/c^2)^{\frac{1}{2}}}-\frac{\bar{p}-\hbar/l}{(m^2+(\bar{p}-\hbar/l)^2/c^2)^{\frac{1}{2}}}. \tag{3.6a}$$

This can again be replaced by $(\hbar/l)(m^2+\hbar^2/l^2c^2)^{-\frac{1}{2}}$.

For given E, the minimum value of a is assumed if the *kinetic energies* of the two particles are of the same order of magnitude. The two terms of (3.6) then become about equal and $l/\lambda\approx(E/(m+E))^{\frac{1}{2}}$. The minimum value of a, as far as order of magnitude is concerned, is again given by (3.2). Similarly, (3.3) also remains valid if one of the two particles has zero rest-mass.

The two-dimensional case becomes simplest if both particles have zero rest-mass. In this case the wave packets do not spread at all and (3.2) can be immediately seen to be valid. In the four-dimensional case, (3.3) again holds. However, its proof by means of explicitly constructed wave packets (rather than reference to the uncertainty relations) is by no means simple. It requires wave packets which are confined in

every direction, do not spread too fast and progress essentially only into one half space (one particle going toward the right, the other toward the left). The construction of such wave packets will not be given in detail. They are necessary to prove (3.2) and (3.3) more rigorously also in the case of finite masses; the preceding proofs, based on the uncertainty relations show only that a and v cannot be *smaller* than the right sides of the corresponding equations. It is clear, in fact, that the limits given by (3.2) and (3.3) would be very difficult to realize, except in the two-dimensional case and for the collision of two particles with zero restmass. In all other cases, the relatively low values of $a_{\min}$ and $V_{\min}$ are predicated on the assumption that the wave packets of the colliding particles are so constituted that they assume a minimum size at the time of the collision. At any rate, (3.2) and (3.3) show that only collisions with a relatively high collision energy, and high energy uncertainty, can be closely localized in space-time.

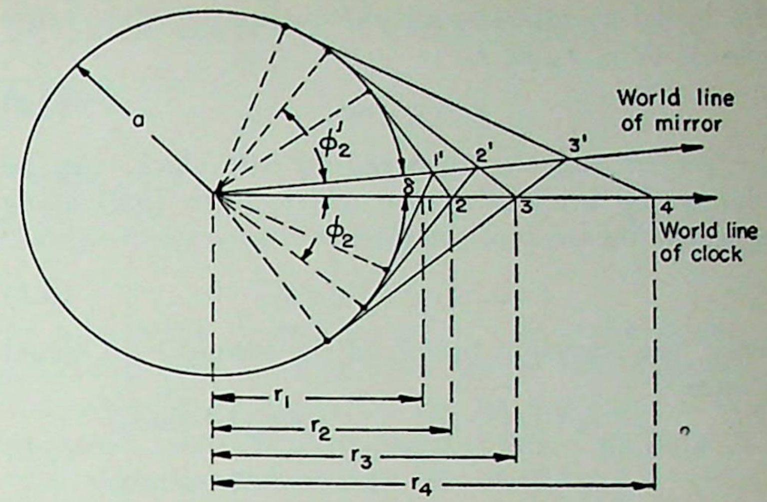


FIG. 7. Analysis of the experiment of Fig. 5. The figure represents a view of the hyperboloid of deSitter space, viewed along its axis. Every point of the plane which is outside the circle corresponds to two points of the deSitter world with one spatial dimension, those with oppositely equal times. The first light signal is emitted at 1, reaches the mirror at 1′, and returns to the clock at 2. The paths of the second and third light signals are 22′3 and 33′4.

APPENDIX IV

Let us denote the components of the vector from event 1 to event 2 by x_i, the components of the unit vector along the world line of the first clock at event 1 by e_i. The components of the first light signal are x_i+te_i, that of the second light signal $x_i-t'e_i$. Hence

$$g^{ik}(x_i+te_i)(x_k+te_k)=0 \quad (4.1)$$

$$g^{ik}(x_i-t'e_i)(x_k-t'e_k)=0. \quad (4.2)$$

Elimination of the linear terms in t and t' by multiplication of (4.1) with t' and (4.2) with t and addition gives

$$2g^{ik}x_ix_k+2tt'g^{ik}e_ie_k=0. \quad (4.3)$$

Since e is a unit vector $g^{ik}e_ie_k=1$ and (4.3) shows that the space-like distance between points 1 and 2 is $(tt')^{\frac{1}{2}}$.

APPENDIX V

Since the measurement of the curvature, described in the text, presupposes *constant curvature* over the space-time domain in which the measurement takes place, we use a space with constant curvature, or, rather, part of a space with constant curvature, to carry out the calculation. We consider only one spatial dimension, i.e., a two-dimensional deSitter space. This will be embedded, in the usual way, in a three-dimensional space[18] with coordinates x, y, τ. The points of the deSitter space then form the hyperboloid

$$x^2+y^2-\tau^2=a^2, \quad (5.1)$$

where a is the "radius of the universe." As coordinates of a point we use x and y, or rather the corresponding polar angles r, ϕ. The metric form in terms of these is

$$ds^2=\frac{a^2}{r^2-a^2}dr^2-r^2d\phi^2. \quad (5.2)$$

Two points of deSitter space correspond to every pair r, ϕ (except $r=a$): those with positive and negative $\tau=(r^2-a^2)^{\frac{1}{2}}$. This will not lead to any confusion as all events take place at positive τ. The null lines (paths of light signals) are the tangents to the $r=a$ circle.

The experiment described in the text can be analyzed by means of Fig. 7. For the sake of simplicity, the clock and mirror are assumed to be "at rest," i.e., their world lines have constant polar angles which will be assumed as 0 and δ, respectively. The first light signal travels from 1 to 1′ and back to 2, the second from 2 to 2′ and back to 3, the third from 3 to 3′ and back to 4. The polar angle of the radius vector which is perpendicular to the first part 22′ of the world line of the second light signal is denoted by ϕ_2. The construction of Fig. 7 shows that angle ϕ_2' which the world line of the mirror includes with the radius vector perpendicular to the second part 2′3 of the second light signal's world line is

$$\phi_2'=\phi_2+\delta. \quad (5.3)$$

The angles ϕ_1, ϕ_1', ϕ_3, ϕ_3' have similar meanings; they are not indicated in the figure in order to avoid overcrowding. For reasons similar to those leading to (5.3), we have

$$\phi_3=\phi_2'+\delta=\phi_2+2\delta \quad (5.3a)$$

$$\phi_1=\phi_2-2\delta \quad (5.3b)$$

$$\phi_4=\phi_3+2\delta=\phi_2+4\delta. \quad (5.3c)$$

[18] See, e.g., H. P. Robertson, Revs. Modern Phys. 5, 62 (1933).

The radial coordinates of the points 1, 2, 3, 4 are denoted by r_1, r_2, r_3, r_4

$$r_i = a/\cos\phi_i. \tag{5.4}$$

The proper time t, registered by the clock, can be obtained by integrating the metric form (5.2) along the world line $\phi=0$ of the clock

$$t = a \ln[r+(r^2-a^2)^{\frac{1}{2}}]. \tag{5.5}$$

Hence, the traveling time t_2 of the second light signal becomes

$$t_2 = a \ln\frac{r_3+(r_3^2-a^2)^{\frac{1}{2}}}{r_2+(r_2^2-a^2)^{\frac{1}{2}}} = a \ln\frac{\cos\phi_2(1+\sin\phi_3)}{\cos\phi_3(1+\sin\phi_3)}. \tag{5.6}$$

Similar expressions apply for the traveling times of the first and third light signals; all ϕ can be expressed by means of (5.3a), (5.3b), (5.3c) in terms of ϕ_2 and δ. This allows the calculation of the expression (3). For small δ, one obtains

$$\frac{t_1-2t_2+t_3}{t_2^2} \approx \frac{11}{a}, \tag{5.7}$$

and Riemann's invariant $R=2/a^2$ is proportional to the square of (5.7). In particular, it vanishes if the expression (3) is zero.

Quantum Theory of Fields and Elementary Particles

W. Heisenberg

Max Planck Institut für Physik, Göttingen, Germany

The present article gives a general discussion of the problems arising in a theory of elementary particles, together with a survey of papers[1-6] that have been published on this subject in German periodicals. These papers deal with a special model for the theory of elementary particles that has been constructed to show some of the main features of such a theory; the author believes that the model does in fact represent a system of elementary particles and their interactions in a manner qualitatively suitable also for the real system of particles.

1. GENERAL REMARKS ON FIELD THEORY AND ELEMENTARY PARTICLES

Any attempt to construct a field theory of elementary particles must from the outset meet the well-known difficulties arising from the combination of the space-time structure of special relativity with quantum theory. Whenever one applies the normal rules of quantization to a differential equation that is Lorentz-invariant and represents interaction between fields, one seems to get divergent results. It may be difficult in the present state of the theory to give a rigorous mathematical proof that these difficulties cannot be completely avoided; but hitherto no solution has been presented. For a time the process of renormalization seemed to offer such a solution. But in the only case where the mathematical structure could be analyzed completely, the Lee model,[7] Källén and Pauli[8] showed that the process of renormalization leads to the implicit introduction of so-called "ghost-states" which destroy the unitarity of the S matrix and thereby violate the rules of quantum theory. To get convergent schemes within the framework of quantum theory one has therefore been forced to introduce the interaction as a nonlocal one,[9] for instance by means of a so-called "cutoff-factor."[10] This however implies deviations from the kind of causality[11] that follows from the space-time structure of special relativity. It is still an open question how serious the deviations from relativistic causality must be in order to ensure a convergent mathematical scheme. But complete local causality seems not to be compatible with quantization. Therefore any field theory of elementary particles or of matter must start by offering some solution to the central mathematical problem: how to combine quantization with a certain greater or lesser degree of relativistic causality. Judging from the experiments, the deviations from causality can scarcely exceed space-time regions of the order of 10^{-13} cm.

The next important problem concerns the fundamental quantities that appear in the mathematical formulation of the theory of elementary particles. Nearly all conventional theories start by introducing some field operators representing the wave fields connected with some specified elementary particles, e.g., meson field or electromagnetic field operators. This procedure, however, requires a definition of the concept "elementary particle." Is there any criterion by which we can distinguish between an elementary particle and a compound system? Is it any more justified to introduce a meson field into the fundamental equations than, e.g., a hydrogen field or an oxygen field?

The author believes that it is essential for any real progress in the theory of matter to recognize that such criteria do not exist.[12] If one asks physicists how they would like to define the nature of an elementary particle as distinct from a compound system, one frequently gets the following answer: an elementary particle is a particle for which one introduces a separate wave function in the fundamental equations. Sometimes one gets the different answer: any particle with spin <1, charge <2, isotopic spin $<\frac{3}{2}$ is elementary; all other particles are compound systems. Obviously both definitions contain elements of complete arbitrariness. Any careful investigation into the properties of atomic particles shows that there may be quantitative differ-

[1] W. Heisenberg, Nachr. Akad. Wiss. Göttingen **1953**, p. 111.
[2] W. Heisenberg, Z. Naturforsch. **9a**, 292 (1954).
[3] Heisenberg, Kortel, and Mitter, Z. Naturforsch. **10a**, 425 (1955).
[4] W. Heisenberg, Z. Physik **144**, 1 (1956).
[5] W. Heisenberg, Nachr. Akad. Wiss. Göttingen **1956**, p. 27.
[6] R. Ascoli and W. Heisenberg, Z. Naturforsch. **12a**, 177 (1957).
[7] T. D. Lee, Phys. Rev. **95**, 1329 (1954).
[8] G. Källén and W. Pauli, Dan. Mat. Fys. Medd. **30**, No. 7 (1955).
[9] Compare H. Yukawa, Phys. Rev. **76**, 300 (1950) and **77**, 219 (1950); P. Kristensen and C. Møller, Dan. Mat. Fys. Medd. **27**, No. 7 (1952); C. Bloch, Dan. Mat. Fys. Medd. **27**, No. 8 (1952).
[10] Compare G. Wataghin, Z. Physik **88**, 92 (1934) and **92**, 547 (1934), and G. Wentzel, Helv. Phys. Acta **13**, 269 (1940).
[11] Compare M. Fierz, Helv. Phys. Acta **23**, 731 (1950), and E. C. G. Stueckelberg and G. Wanders, Helv. Phys. Acta **27**, 667 (1954).
[12] Compare W. Heisenberg, Naturwissenschaften **42**, 637 (1955).

ences between different particles which suggest that it might be convenient in a given experiment to call one particle elementary and the other a compound system; but no qualitative distinction between elementary and compound can be made. A proton certainly looks like an elementary particle for energies <100 Mev, but it may be considered as composed of a Λ^0 particle and a θ^+ particle in collisions of much higher energies. One might argue that the Λ^0 particle and the θ^+ particle are unstable while the proton is stable, that therefore the proton cannot be composed of Λ^0 and θ^+. The fallacy of the argument is, however, seen at once from the case of the deuteron, which is stable but usually considered as composed of proton and neutron, the latter being unstable. For an understanding of matter and of the atomic particles it is essential to realize that the question whether the proton is elementary or a compound system has no answer. This result should not prevent us from using the term "elementary particle" whenever it is convenient to disregard its inner structure. But it should not be misunderstood as making some specific statement about the nature of the particle.

To avoid the two fundamental difficulties which have been put here at the beginning of the discussion, the efforts of many physicists have in recent years been concentrated on the S matrix.[13] The S matrix is the quantity immediately given by the experiments. It is not difficult to construct S matrices which fulfill the requirements of Lorentz-invariance and unitarity without encountering any divergent terms. At the same time the problem of the "elementary" particles does not occur immediately in the S matrix, since any incoming or outgoing particles are here characterized by their wavefunctions ψ_{in} or ψ_{out}, irrespective of whether they are compound or not.

The S-matrix formalism does not by itself guarantee the requirements of relativistic causality. Therefore many recent investigations have dealt with supplementary conditions to be put on the S matrix to ensure relativistic causality. The best known example is the treatment of the dispersion relations.[14] Insofar as these conditions state relations between observable quantities they will serve as a very useful tool for the interpretation of the experiments. If, however, their mathematical formulation makes use of an extrapolation of the S matrix into regions of momentum space, where the relations $\mathbf{p}^2+\kappa^2=0$ for the respective particles are not fulfilled, their value might be very limited, since it is difficult to see what the momentum vector p_μ of, say, a hydrogen atom in its normal state can mean when the relation $\mathbf{p}^2+\kappa^2=0$ (where κ is the total mass of the atom in its normal state) is not fulfilled. In other words: the extrapolation into these regions of momentum space requires eventually the distinction between elementary particles and compound systems which cannot be more than a more or less suitable convention. It is perhaps not exaggerated to say that the study of the S matrix is a very useful method for deriving relevant results for collision processes by going around the fundamental problems. But these problems must be solved some day and one will then have to look for a mathematical formalism that allows one to calculate the masses of the particles and the S matrix at the same time. The S matrix is an important but very complicated mathematical quantity that should be derived from the fundamental field equations; but it can scarcely serve for formulating these equations.

As a result of the foregoing discussion we can try to state some general principles for a theory which attacks the problem of the fundamental field equations for matter.

1. The field operators necessary for formulating the equations shall not refer to any specified particle like proton, meson, etc.; they shall simply refer to matter in general.

2. The particles (elementary or compound) should be derived as eigensolutions of the field equations.

3. The fundamental field equations must be nonlinear in order to represent interaction. The masses of the particles should be a consequence of this interaction. Therefore the concept of a "bare particle" has no meaning.

4. Selection rules for creation and decay of particles should follow from symmetry properties of the fundamental equations. Therefore the empirical selection rules should provide the most detailed information on the structure of the equations.

5. Besides the selection rules and the invariance properties, the only other guiding principle available seems to be the simplicity of the equations.

The empirical spectrum of elementary particles looks very complex. All hitherto observed particles have lifetimes $>10^{-15}$ sec; for most of them the lifetime is $>10^{-11}$ sec. The natural lifetime for elementary particles that can decay into others would, however, be of the order 10^{-22} to 10^{-23} sec. This comparison shows that the observed particles represent those rare cases where the selection rules provide an exceptionally long lifetime. When one compares in the optical spectra of atoms the number of levels with the natural lifetime of the order 10^{-8} sec with the small number of metastable levels where the lifetime is longer, say by a factor of 10^8, one gets an idea of how complicated the real spectrum of elementary particles may possibly be.

Under these circumstances it seems advisable first to study a simplified model, which may be constructed in accordance with the foregoing principles. The model to be discussed in the following sections is certainly

[13] Compare J. A. Wheeler, Phys. Rev. 52, 1107 (1937), and W. Heisenberg, Z. Physik 120, 513, 673 (1943); Z. Naturforsch. 1, 608 (1946); C. Möller, Kgl. Danske Videnskab. Selskab. 23, No. 1 (1945); 24, No. 19 (1946).

[14] Compare Gell-Mann, Goldberger, and Thirring, Phys. Rev. 95, 1612 (1954).

too simple to represent the real spectrum of elementary particles. But it shows the main features of an adequate theory inasmuch as it represents a world consisting of elementary particles with qualitatively similar properties to our own.

2. MODEL FOR A THEORY OF MATTER

(a) Wave Equation and Quantization

The model which has been studied in detail in the series of papers mentioned above[1–6] starts from the equation*

$$\gamma_\nu \frac{\partial \psi}{\partial x_\nu} - l^2 \psi(\psi^+ \psi) = 0. \quad (1)$$

$\psi(x)$ is defined as a spinor wave function representing matter. The equation has been chosen according to the principles 1, 3, 5 of the foregoing section. Equation (1) is certainly too simple to represent the real particles since it does not contain the isobaric spin variable.

Equation (1) defines a quantum theory for matter only if commutation relations for the operator $\psi(x)$ are added. At this point one meets the difficulty mentioned in the beginning. This difficulty will be discussed in some detail.

To get some information on the possible assumptions for the commutation relations and on the eventual connection between "commutator" and "propagator," it is convenient to consider the operator:

$$\chi_\alpha(x,x') = \exp\{-i[a_\nu \psi_\nu^+(x') + \text{conj.}]\}\psi_\alpha(x) \times \exp\{+i[a_\nu \psi_\nu^+(x') + \text{conj.}]\}. \quad (2)$$

("Conj." in the exponent means the Hermitian conjugate.) a_ν is an arbitrary constant spinor that anticommutes with $\psi(x)$ and $\psi^+(x)$ everywhere. $\chi_\alpha(x,x')$, considered as a function of x, satisfies the wave Eq. (1). If the components a_ν are chosen as very small, $\chi_\alpha(x,x')$ can be expanded as

$$\chi_\alpha(x,x') = \psi_\alpha(x) - ia_\nu[\psi_\alpha(x)\psi_\nu^+(x') + \psi_\nu^+(x')\psi_\alpha(x)] - ia_\nu^*[\psi_\alpha(x)(\psi_\nu^+(x'))^* + (\psi_\nu^+(x'))^*\psi_\alpha(x)] + \cdots. \quad (3)$$

(The star * denotes Hermitian conjugation.) To preserve relativistic causality, one usually assumes the anticommutators to be zero for space-like distances $(x-x')^2 > 0$. This assumption means that $\chi_\alpha(x,x')$ corresponds to a solution of (1), in which a secondary wave starts from the singular point $x = x'$ and fills the cones of future and past. The amplitude of this wave will be very small everywhere except in the immediate neighborhood of the lightcone, if the a_ν are very small. In all conventional forms of quantum theory it has been assumed that the anticommutators (or commutators) are singular c-number functions in the immediate neighborhood of the lightcone. If this assumption is taken over into Eqs. (2) and (3), one can put

$$\chi_\alpha(x,x') = \chi_\alpha^0(x,x') + c_\alpha(x-x'), \quad (4)$$

where $c_\alpha(x-x')$ is a c number function (except for a sign function occurring in the a_ν) and contains the singularities at the lightcone, while $\chi_\alpha^0(x,x')$ is smooth at the light cone. One may further assume in defining $c_\alpha(x-x')$ that the vacuum expectation value of $\chi_\alpha^0(xx')$ vanishes. By inserting (4) into (1) and taking the vacuum expectation value, one finds for c_α as function of x an equation of the type

$$\gamma_\mu \frac{\partial c}{\partial x\mu} - l^2 c(c^+ c) - \kappa(s)\cdot c = l^2 \langle \Omega | \chi^0(\chi^{0+}\chi^0) | \Omega \rangle, \quad (5)$$

where $s = (x-x')^2$, and $\kappa(s)$ is essentially given by the vacuum expectation value of $|\chi^0(x,x')|^2$. The right side of (5) vanishes in the neighborhood of the lightcone.

To study the possible assumptions for the anticommutator

$$S_{\alpha\nu}(xx') = \psi_\alpha(x)\psi_\nu^+(x') + \psi_\nu^+(x')\psi_\alpha(x), \quad (6)$$

one has therefore to look for the solutions of the classical nonlinear wave equation (5) where the right-hand side can be put equal to zero, since we are only interested in the neighborhood of the lightcone. In this region $S_{\alpha\nu}(x-x')$ should behave as $c_\alpha(x-x')$ for small a_ν; or more accurately: one should study the continuous group of solutions of (5) that corresponds to different values of a_ν and should finally assume

$$S_{\alpha\nu}(x-x') \sim -\lim_{a_\nu \to 0} \left(\frac{\partial c_\alpha(x-x')}{\partial a_\nu} \right) \quad (7a)$$

near the lightcone. For the second anticommutator it is plausible to assume as usual

$$\psi_\alpha(x)\psi_\nu(x') + \psi_\nu(x')\psi_\alpha(x) = 0 \quad (7b)$$

near the lightcone.

For the definition of a quantum theory for matter it will be sufficient to state, besides (1), the commutation rules only near the lightcone, since their value in other parts of space and time will then follow from (1). The equations (5) and (7a) state the required connection between "anticommutator" and "propagator."

Mathematical analysis of (5) shows that its solutions exhibit infinite oscillations in the neighborhood of the light cone.[2] This is entirely different from the behavior of $S_{\alpha\nu}$ in a linear wave theory, where $S_{\alpha\nu}(x-x')$ behaves like the derivative of the Dirac δ function.

At this point we meet the difficulty mentioned in the beginning of Sec. 1. If we are primarily interested

* The original papers start from an equation that differs from Eq. (1) by the sign of the second term. But—as has been pointed out[6]—if one wants to use conventional formalism, one should say that the calculations actually refer to Eq. (1).

in the vacuum expectation value

$$S_{\alpha\nu}(xx') = \langle\Omega|\psi_\alpha(x)\psi_\nu^+(x') + \psi_\nu^+(x')\psi_\alpha(x)|\Omega\rangle$$
$$= \sum_\Phi \langle\Omega|\psi_\alpha(x)|\Phi\rangle\langle\Phi|\psi_\nu^+(x')|\Omega\rangle$$
$$+ \sum_\Phi \langle\Omega|\psi_\nu^+(x')|\Phi\rangle\langle\Phi|\psi_\alpha(x)|\Omega\rangle, \quad (8)$$

each group of intermediate states Φ, belonging to a definite mass eigenvalue κ, contributes an ordinary Schwinger function $S_{\alpha\nu}^\kappa(xx')$ belonging to that mass; and the δ and δ' functions of the different intermediate masses all add up with the same sign. Therefore the sum over Φ in (8) can never lead to the behavior of $c_\alpha(x-x')$, the infinite oscillations near the light cone, since the integral contribution of the oscillations vanishes near the lightcone, while that of the δ functions does not vanish.

At a finite distance from the lightcone the sum over Φ in (8) can very well represent a solution of (5) for small a_ν, since here the nonlinear term $c(c^+c)$ can be neglected and the variable coefficient $\kappa(s)$ in (5) corresponds to the different mass values of the states Φ. But near the light cone the rules of quantization must be changed in order to avoid the contradiction between (5) and (8).

(b) Hilbert Space II and the Unitarity of the *S* Matrix

The only feasible way of getting rid of the δ and δ' functions on the lightcone in (8) seems to be an extension of Hilbert space by the introduction of new intermediate states Φ, which change the metric of Hilbert space into an indefinite one. This extra group of states, called Hilbert space II in the papers mentioned, is chosen so as to cancel the δ and δ' functions on the lightcone in (8). They do not contribute to $S(xx')$ in other parts of space-time. But they will then contribute to the functions $S_1(xx')$ or $S_F(xx')$ also in other parts of space-time, if the normal Schwinger relation

$$S_1(\mathbf{r},t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} S(\mathbf{r},t-t')dt'/t' \quad (9)$$

is taken over from conventional theory.

If, in a first approximation, one considers in Hilbert space I (which comprises all physical states of the system) only one group of intermediate states Φ, belonging to a single mass κ, the corresponding part of the S function in momentum space would be

$$\tfrac{1}{2}S(p) = \frac{p_\mu\gamma_\mu + i\kappa}{\mathbf{p}^2+\kappa^2}. \quad (10)$$

The subtraction of the δ and δ' functions on the light cone can then be performed by putting, in this same approximation,

$$\tfrac{1}{2}S(p) = \frac{p_\mu\gamma_\mu + i\kappa}{\mathbf{p}^2+\kappa^2} - \frac{p_\mu\gamma_\mu + i\kappa}{\mathbf{p}^2} + \frac{p_\mu\gamma_\mu\kappa^2}{(\mathbf{p}^2)^2}. \quad (11)$$

The contributions from Hilbert space II appear to belong to a mass value 0. But by writing the contributions in a different way, namely,

$$\frac{\kappa^2 p_\mu\gamma_\mu}{(\mathbf{p}^2)^2} - \frac{p_\mu\gamma_\mu + i\kappa}{\mathbf{p}^2} = \lim_{\epsilon\to 0}\frac{p_\mu\gamma_\mu\cdot\kappa^2}{\epsilon}\left(\frac{1}{\mathbf{p}^2} - \frac{1}{\mathbf{p}^2+\epsilon}\right)$$
$$\times\left(1 - \frac{ip_\mu\gamma_\mu}{\kappa} - \frac{\mathbf{p}^2}{\kappa^2}\right), \quad (12)$$

one sees that the extra states of Hilbert space II can be considered as "dipole" states, composed of one normal state with mass 0 and a "ghost state" with mass $\sqrt{\epsilon}\to 0$. ("Ghost state" on account of the negative sign in the metric.) As a result of the dipole character of these states, their representatives do not depend simply exponentially on space-time as do the representatives of states in Hilbert state I. A simple calculation of the representatives on the basis of (11) and (12) shows that either the covariant or the contravariant representative has the space-time dependence (in a suitable coordinate system)

$$(t + \text{const})e^{i\mathbf{px}} \quad (13)$$

while the other representative has the usual exponential form $e^{i\mathbf{px}}$.

This is an important result, because it shows that the states of Hilbert space II—contrary to the "ghost states" of Källén and Pauli[8]—do not destroy the unitarity of the S matrix. In fact if in a collision problem all incoming waves belong to Hilbert space I, the total wave function depends exponentially on space-time in both representations. This behavior cannot be changed by collision [on account of the invariance of Eq. (1) for the inhomogeneous Lorentz group], therefore also the outgoing waves cannot contain contributions from Hilbert space II, since they would destroy the exponential form of the total wave function. Therefore the extension of Hilbert space, which was necessary in order to avoid contradiction between (5) and (8), does not destroy the unitarity of the S matrix and may therefore be compatible with the experimental results. (Introduction of Hilbert space II has in this respect some resemblance to the method of Gupta[15] in quantum electrodynamics.) Of course this extension does mean a certain deviation from local relativistic causality. The local behavior of $\psi(x)$ cannot be interpreted in the usual manner, since $\psi(x)$ contains contributions from Hilbert space II, and their conven-

[15] S. N. Gupta, Proc. Roy. Soc. (London) 63, 681 (1950); 64, 850 (1951); K. Bleuler, Helv. Phys. Acta 23, 567 (1950). Compare P. A. M. Dirac, Proc. Roy. Soc. (London) A180, 1, (1942).

tional interpretation would, on account of the negative sign in the metric, lead to negative probabilities, which have no meaning.†

Actually these contributions from Hilbert space II change the form of $S_1(xx')$ and $S_F(xx')$ fundamentally. For large values of $(x-x')^2>0$ these functions decrease more slowly (like $(x-x')_\nu\gamma_\nu/(x-x')^2$) than the usual Schwinger functions and this behavior leads to long range forces between the particles. It will be shown later that it is through these contributions that Eq. (1) contains quantum electrodynamics with a specific value of the Sommerfeld fine-structure constant.

(c) Methods of Integration

The equations (1) and (11) should be sufficient to define a quantum theory for matter. Since Eq. (11) is only a first approximation, it should generally be replaced by its precise form:

$$\tfrac{1}{2}S(p)=\int\rho(\kappa)d\kappa\left[\frac{p_\mu\gamma_\mu+i\kappa}{\mathbf{p}^2+\kappa^2}-\frac{p_\mu\gamma_\mu+i\kappa}{\mathbf{p}^2}+\frac{p_\mu\gamma_\mu\kappa^2}{(\mathbf{p}^2)^2}\right] \quad (14a)$$

with

$$\int\rho(\kappa)d\kappa=1,$$

where $\rho(\kappa)$ is the mass spectrum of fermions. $\rho(\kappa)$ should be derived from Eqs. (1) and (14).

On account of the infinite oscillations on the light-cone, one can simply put

$$S_{\alpha\nu}(xx')=0 \quad \text{for} \quad (x-x')^2=0. \qquad (14b)$$

The only method of integration that has been used so far for Eqs. (1) and (14) is what is sometimes called the new Tamm-Dancoff method.[16] One considers τ functions of the general type

$$\tau_{\alpha\beta\gamma}(x_1x_2|x_3)=\langle\Omega|T\psi_\alpha(x_1)\psi_\beta(x_2)\psi_\gamma^+(x_3)|\Phi\rangle \quad (15)$$

as covariant representatives of the states Φ. (T means the time-ordered product.) By means of Eq. (1) one can find differential equations, which connect the derivatives of one τ function with another τ function with a number of variables larger by two. This differential equation can be integrated by means of the Green function $G_F(xx')$ of the Dirac equation for mass zero. (The Feynman functions G_F are chosen in order to satisfy the boundary conditions.) By repeating this process a connection can be established between the original τ function and another one, where the number of variables is larger by any even number. This latter τ function is then expressed in terms of the so-called φ functions through the process of contraction:

$$\tau(x_1x_2\cdots|y_1y_2\cdots)=\varphi(x_1x_2\cdots|y_1y_2\cdots)-\tfrac{1}{2}S_F(x_1y_1)\times\varphi(x_2\cdots|y_2\cdots)-\cdots+\tfrac{1}{4}S_F(x_1y_1)S_F(x_2y_2)\times\varphi(x_3\cdots|y_3\cdots)+\cdots-\cdots, \quad (16)$$

and finally all those φ's are neglected, the number of variables of which is larger than that of the original τ function.

In this way a linear integral equation for the τ function is established that can be used for the determination of the mass eigenvalues or the S matrix. This integral equation can be represented by a Feynman graph, consisting of two types of lines, one representing the G_F function, the other representing the S_F function. From (1) and (16) one can easily derive the following rules for these graphs:

1. In every point of the graph four lines meet. The initial and the final points of the graph are considered as identical, or the graph is repeated as a periodic pattern.
2. The numbers of S_F and G_F lines in the graph are equal.
3. Every point is connected with one of the final points through one sequence of G lines.
4. The connection of two lines at a point means matrix multiplication of the respective operators.
5. The kernel of the integral equation is given, when the points of the graph and the pattern of G lines are given. One has to sum over the contributions from all possible S line patterns belonging to the given G line pattern.

This scheme differs from conventional formalism by two characteristic features. The contraction is performed by the function S_F, which is *not* identical here with the Green function G_F; and there are no δ terms in the equations, which in the usual theory arise from exchange of ψ-factors $\psi(x_1)$ and $\psi^+(x_2)$ at the point $t_1=t_2$. Here all $\psi(x)$ functions anticommute on a given subspace $t=$const. Therefore, quantization is introduced only by the contractions.

This last difference is a significant consequence of the nonlinearity of the equations. If in a linear theory the commutator vanished on a given subspace $t=$const, it would, by virtue of the wave equation, vanish everywhere. Therefore one must start with a δ function at the point $x=x'$ on the subspace $t=$const. In a nonlinear theory the commutator can vanish everywhere on $t=$const and still be different from zero at other times. It is a well-known property of nonlinear equations—the fact that a solution sometimes cannot be continued—which has to be used here in the definition of quantization.

† *Note added in proof.*—A mathematical analysis of this method of quantitization can be obtained from the Lee model. The constants g_0 and m_v of the Lee model can be adjusted to let the energies of the normal V particle and the "ghost-state" coincide. In this case the two states form a dipole as in (12), the renormalized wave functions ψ_v commute on the subspace $t=$const and the two parts of Hilbert space are distinguished by the asymptotic behavior of the wave functions. The details will be published in *Nuclear Physics*.

[16] M. Gell-Mann and F. Low, Phys. Rev. 84, 350 (1951); Freese, Z. Naturforsch. 8a, 776 (1953).

The contraction function S_F can, in a low approximation, be represented by (11), and the eigenvalue κ is to be obtained from the integral equation. In higher approximations one may consider several eigenvalues or, finally, the form (14).

Whether this whole procedure will, by going from low to high approximations, converge to a final solution, is still an open question. It does give finite results in any approximation. But it may be necessary to define the integral equations by averaging over certain groups of G-line graphs, in order to obtain convergence. This problem has been treated in detail by Matthews and Salam for conventional formalism.[17] It may also be necessary to replace the φ functions by slightly different groups of functions, as suggested by Maki.[18] The procedure has been studied in detail in the example of the anharmonic oscillator. But whatever the results of such mathematical investigations will be, one will probably get some—though inaccurate—information on the solutions of (1) and (14) by using the rules described in this section for the low approximations.

(d) The Lowest Eigenvalues

The lowest eigenvalues of (1) and (14) have been calculated along the lines described in the foregoing section.[3] For the eigenvalues of the fermion type all graphs of the type of Fig. 1 have been combined to



FIG. 1.

define the integral equation. (The G lines are given as full, the S lines as dotted lines.) Only τ functions with one or three variables have been used.

The result was that there exists—in this approximation—only one eigenvalue for the mass κ of a fermion:

$$\kappa = 7.426/l. \tag{17}$$

The spin of this particle is 1/2.

The bosons have been calculated from the graph given in Fig. 2. Only τ functions for two variables $[\psi_\alpha(x)$



FIG. 2.

and $\psi_\beta^+(x)$ with equal space-time coordinates$]$ have been used. Four different bosons with nonvanishing masses were found, with mass values, spins, and parities as shown in Table I.

TABLE I.

Mass κl	Spin	Parity
0.33	1	−1
0.95	0	1
1.74	0	−1
3.32	0	1

It is interesting that (1) and (14) give for the boson masses values which are considerably smaller than the mass of the fermion. This fits well with the empirical fact that the masses of π meson and θ meson are considerably smaller than the mass of the nucleon.

It may be useful at this point to compare the properties of the model, given by (1) and (14), with the general principles laid down at the end of Sec. 1.

The wave function $\psi(x)$ in (1) refers to matter in general, not to a specified particle. The particles do actually come out as the eigensolutions of the equations. Since $\psi(x)$ has been chosen as spinor, while the commutation relation states the value of the anticommutator, all particles with half quantum spin obey Fermi statistics, all particles with integer spin obey Bose statistics. States of the first kind are obtained when one applies an odd number of $\psi(x)$ operators on the vacuum state, the Bose states are created by applying an even number of ψ's.

The nonlinear term in (1) is multiplied with a coupling constant with the dimension of a length. A variation of l will simply change all mass eigenvalues by a constant factor. The ratio of the eigenvalues depends only on the general form of the nonlinear term. The masses of the particles are entirely a product of interaction, namely of the nonlinear term. Therefore their interactions are given simultaneously with their masses; the concept of a bare particle has no meaning in this theory.

Besides the bosons given in Table I analysis of (1) reveals the existence of still another group of boson states belonging to the rest mass 0 (which had not been expected in the papers[1,2]). Their existence can be seen as follows.[3] When one applies the integral operator represented by the graph of Fig. 2 on a τ function belonging to a total momentum vector J_μ with $J_\mu^2=0$, one gets generally an infinite result. But there are special forms of the τ function

$$\tau_{\mu\beta}(x|x) = \langle\Omega|\psi_\alpha(x)\psi_\beta^+(x)|\Phi\rangle = e^{iJ_\mu x_\mu}\tau_{\alpha\beta}, \tag{18}$$

for which the divergencies disappear. These special τ functions can be used to satisfy the integral equations.

As conditions for the cancellation of the infinite terms one finds

$$\begin{aligned} J_\mu\gamma_\mu(\tau - \text{spur}\,\tau) &= 0, \\ (\tau - \text{spur}\,\tau)J_\mu\gamma_\mu &= 0, \\ \gamma_\mu\tau\gamma_\mu &= 0. \end{aligned} \tag{19}$$

[17] P. T. Matthews and A. Salam, Proc. Roy. Soc. (London) A221, 128 (1954).
[18] Z. Maki, Progr. Theoret. Phys. 15, 237 (1956).

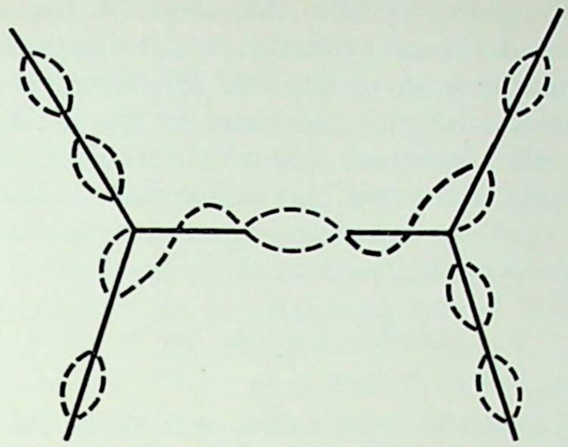


FIG. 3.

If one introduces two four-vectors A_ν, which satisfy $A_\mu J_\mu = 0$ and are linearly independant of each other and of J_μ (there are just two such vectors), one can solve the equations (19) for τ by putting

$$\tau = J_\mu A_\nu \gamma_{\mu\nu} \tag{20}$$

where

$$\gamma_{\mu\nu} = (i/2)(\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu). \tag{21}$$

There are two independent solutions, since there are two independent vectors A_ν. Another equivalent solution would be

$$\tau = J_\mu A_\nu \gamma_{\mu\nu} \gamma_5 \tag{22}$$

and actually this solution has been discussed.[3] But it is easily seen that the solutions (22) are not linearly independent of (20); they can be obtained from (20) by a linear transformation. Therefore there are just two independent solutions for each vector J, characterized by the "vectors of polarization" A_ν. It has been shown[3] that these bosons of rest mass 0 have all the transformation properties of light quanta and belong to the spin spin values ± 1 (the axis of the spin being parallel or antiparallel to the direction of propagation).

Existence of these bosons is closely connected with the existence of long range interactions between particles, which in turn are a consequence of the dipole states in Hilbert space II.

(e) Interaction between Particles. Electrodynamics and the Fine Structure Constant

The interaction between particles in collision processes can be treated by essentially the same method as the mass eigenvalues. To give a graph-picture of the interaction one can represent the incoming and outgoing particles by infinite tails, which are just periodic repetitions of the elementary graphs used for the calculation of their masses.[3] The interaction is then represented by a pattern of S and G lines connecting the periodic tails. A typical example is shown in Fig. 3.

It follows from rule 3 in section 2(c) for the graph patterns that there must be a break in the G lines connecting the two particles, which can be filled by two S lines, as is shown in Fig. 3. The two S_F functions vary for large space-like distances of x and x' as $(x-x')_\nu \gamma_\nu/(x-x')^2$, their product varies roughly as $(x-x')^{-2}$. Therefore the two S functions, connecting the inner endpoints of the G lines, show together a behavior similar to that of the D_F function of quantum electrodynamics, and produce long range forces between the particles qualitatively similar to Coulomb forces.

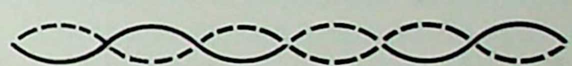
Of course the multitude of graphs of the general type of Fig. 3 contains many different interactions, produced by all the different fields that belong to the different particles, bosons, and fermions. Actually the interaction will be a mixture of contributions from all particles that cannot be disentangled. Only in the case of bosons of rest-mass 0 can one separate their influence by studying the long range forces, since all other particles would produce only short range forces.

Calculation of the exact form of the long range forces is, however, rather complicated. One may for instance try to derive the cross section for collisions in which very little momentum J_μ is transferred from one fermion to the other:

$$P_\mu^{(1)} - P_\mu^{(1)\prime} = J_\mu = P_\mu^{(2)\prime} - P_\mu^{(2)}; \quad |J_\mu^2| \ll \kappa^2. \tag{23}$$

In this case it is necessary to take into account very long tails of the type of Fig. 4 in between the fermions,

FIG. 4.



because for very small J_μ the tail means bosons (of rest-mass 0) that are nearly free; and for free bosons the tail could be infinitely long. Therefore the calculation has been carried out in two steps. In the paper cited in reference 3, the summation over all tails of different lengths was performed and led to an operator connecting the two fermions, which was composed of the eigenfunctions (20) or (22) of the light quanta [(104) of the paper cited], thereby showing that the long range forces are actually transferred by means of the field corresponding to the light quanta of Sec. 2(d). Then the operator O^S (page 440 of the paper), which connects as a kind of vertex part the former operator with either of the fermions, had to be calculated. O^S is represented by a graph of the general type of

FIG. 5.

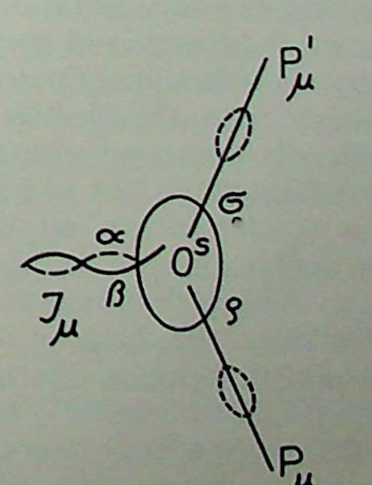


Fig. 5. From arguments of symmetry and invariance[6] O^S has the general form

$$O_{\alpha\beta|\rho\sigma}{}^{\nu}=\text{const}\frac{J_\mu}{\sqrt{J^2}}(\gamma_{\mu\nu})_{\alpha\beta}(\gamma_\nu)_{\rho\sigma}. \tag{24}$$

The constant factor could be calculated only rather inaccurately, since its determination required the use of special methods of approximation the accuracy of which is somewhat doubtful. Finally the transition matrix element for the collision turns out to be

$$\frac{\text{const}}{J^2}[u^+(P^{(1)})\gamma_\mu u(P^{(1)})][u^+(P^{(2)})\gamma_\mu u(P^{(2)})]$$

$$\cdot\delta(P^{(1)}-P^{(1)\prime}+P^{(2)}-P^{(2)\prime})-\text{exchange term}. \tag{25}$$

[$u(P)$ is the eigenfunction belonging to a fermion of momentum vector P.] This is exactly the form for the transition element in quantum electrodynamics. The value of the fine structure constant α_F was determined [using the inaccurate value of the constant in (24)] with the result

$$\alpha_F\approx 1/267. \tag{26}$$

This shows that quantum electrodynamics with a special value of the fine structure constant is contained in the Eqs. (1) and (14) of our model for a theory of matter. One could not expect that the value of the fine structure constant should just be the empirical value 1/137, since the model theory is not yet the correct theory. But the fact that one gets a definite value of this constant of the right order of magnitude, seems to indicate that the model is in this respect not too far from the truth.

Equation (25) is perhaps not sufficient to show that (1) and (14) contain the complete scheme of quantum electrodynamics. Actually (25) has been supplemented[6] by proof that the conservation of charge holds generally and that one can construct field operators $F_{\mu\nu}$, which obey the Maxwell equations. These field operators are closely connected with the operators $\psi(x)\gamma_{\mu\nu}\psi^+(x)$ which have the same transformation properties under the Lorentz group, but they are not identical with them. The operators $\psi(x)\gamma_{\mu\nu}\psi^+(x)$ represent, besides the fields $F_{\mu\nu}$, other properties of matter, for instance the spin density of the fermions. Conservation of charge follows essentially from the identity $\partial^2 F_{\mu\nu}/\partial x_\mu \partial x_\nu=0$ and is not primarily connected with the invariance of the fundamental equations (1) and (14) against the transformation $\psi\rightarrow\psi e^{i\alpha}$. Since all particles can be considered as combinations of fermions, their charge can only be an integral multiple of the elementary charge of the fermion. The charge of the antifermion is opposite to that of the fermion. The bosons calculated from the graph [Fig. 2 in Sec. 2(d)] are neutral.

A few remarks may be added concerning the smallness of the fine structure constant. This smallness certainly has nothing to do with the value of the coupling parameter l. It is primarily a consequence of the mathematical form of the two functions G_F and S_F which are combined in the graph of Fig. 2. Their form leads to the conditions (19) which are very restrictive with respect to the possible solutions for $\tau_{\alpha\beta}$. In calculating the operator connecting the two fermions one has to make use of the relation:

$$\delta_{\alpha\beta}\delta_{\rho\sigma}=\tfrac{1}{4}\sum_N \gamma_{\alpha\sigma}{}^N\gamma_{\beta\rho}{}^N, \tag{27}$$

where the sum is to be taken over all 16 elements of the Dirac algebra. Of this sum only the tensor terms $(\gamma_{\mu\nu})_{\alpha\sigma}(\gamma_{\mu\nu})_{\beta\rho}$ contribute to the interaction on account of the conditions (19), and only three of the six tensor terms contribute to the electromagnetic forces on account of the properties of the operator O^S in (24). These reasons for the smallness of the fine structure constant would therefore remain even if the form of the nonlinear term in (1) were altered.

If one were to define a corresponding coupling constant for the short range forces, its value should be of the order unity, since there are many bosons of different symmetries and the restrictive conditions (19) do not appear. This result fits well with the empirical fact of a large coupling constant for the Yukawa interaction. But one should remember that it should not, according to the graphs of the type of Fig. 3, be possible to single out a special short range interaction for a given boson from the rest of interactions.

(f) Nonlinear Integral Equations for S_F

The form of the function S_F has so far been derived from the assumed existence of fermions together with the assumption of the states of Hilbert space II. Since S_F is identical with a τ function of two variables it should also be possible to derive S_F from integral equations in the same manner as is done with the other τ functions. There is only the one essential difference

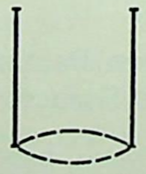


FIG. 6.

that the integral equations for S_F would be nonlinear, since S_F enters twice into these equations, once as the initial τ function and again as the function of contraction. For instance, a graph of the type of Fig. 6 would lead to an integral equation of the general type

$$\gamma_\mu\frac{\partial S_F}{\partial x_\mu}\sim G_F S_F{}^3. \tag{28}$$

An integral equation of this type would probably, on account of its similarity to (1), lead to oscillations near the light cone, which do not show up in the

approximate solution (11). Therefore one could not expect to derive the solution (11) from an approximate integral equation like (28); still the exact form (14) could quite well be a solution of the exact integral equations. Actually it should be possible, at least in principle, to determine the spectrum $\rho(\kappa)$ of the fermions from this integral equation.

An attempt[3] was made to use differential equations corresponding to the integral equations of Fig. 6 for a determination of the asymptotic behavior of S_F at large space-like values of x and x'. There were two differential equations, one connecting $S_F(xx')$ with $\tau(xx|xx')$, the other connecting $\tau(xx|xx')$ with $\tau(xxx'|xx'x')$ and by means of three contractions again with $S_F(xx')$. Kita[19] has pointed out that these equations[3] contained an error of sign and that, after the correction of the error, the equations do not lead to the correct asymptotic behavior of S_F. But a closer investigation has since shown that actually the whole procedure for the calculation cannot be justified, since $\tau(xx|xx')$ is identical with $\varphi(xx|xx')$, on account of (14b), $S_F(0)=0$. The evaluation of $\tau(xxx'|xx'x')$ by means of contractions, however, is only possible if the φ functions of four variables can be neglected. The second differential equation would therefore use the φ function of four variables on the left-hand side while it would neglect it on the right-hand side. This cannot lead to any reasonable approximation.

It has so far not been possible to improve this calculation and to get information on the form of the S_F function from the nonlinear integral equations, since the construction of the φ functions with four variables would require very complicated mathematical investigations.

3. EXTENSION OF THE MODEL TOWARDS A REALISTIC THEORY OF MATTER

(a) Introduction of the Isobaric Spin

The model of (1) and (14) cannot represent the real system of elementary particles since the isobaric spin has not yet been introduced into the equations. (One could of course argue that the isobaric spin should not be put into the equations, but that it should come out of them, since the isobaric spin is closely connected with the charge. It may be that a careful study of the way in which the charge is attached to the particles will lead to a deeper insight into the nature of isobaric spin. For the time being however it seems necessary to introduce the isobaric spin into the equations.) To extend the model by its introduction to a more realistic one, it will not be sufficient to define the wave function ψ as spinor both in ordinary space and in isobaric spin space; for then all particles with half quantum spin would also have half-quantum isobaric spin, and integer spin values would be connected with integer isobaric spin. But the selection rules put forward by Pais,[20] Gell-Mann,[21] Nakano and Nishijima[22] and others seem to show that, e.g., the Λ^0 particle has the spin 1/2 and isobaric spin 0. Therefore one needs at least two fundamental fields (as has been suggested by Goldhaber[23]), say ψ and χ, the one of which is a spinor in ordinary and in iso-space, while the other, χ, may be a spinor in ordinary space, but a scalar in iso-space.

One may of course choose other combinations for the two fields, but these assumptions are the simplest ones and allow one to construct a theory very similar to the model (1) and (14). We have now to investigate how far one can come with two such functions in describing the real system of elementary particles.

In the system of the real particles one usually distinguishes three types of interactions: the strong interaction, for instance between nucleons, π mesons, and hyperons; the electromagnetic interaction; and the weak interaction which for instance produces the decay of the Λ^0 particle or other radioactive processes.

The coupling constant for the weak interactions is extremely small compared to normal nuclear constants of the same dimension. They will certainly not play any appreciable role in determination of the masses. Therefore it will be convenient first to neglect the weak interactions all together and to introduce them later as a very small perturbation.

The electromagnetic interaction cannot, if we follow the model of (1) and (14), be separated from the strong interaction, since it is one of its consequences. On the other hand from the empirical selection rules one is inclined to think that the strong interactions are invariant against any rotations in isobaric spin space, while the electromagnetic interaction is only invariant against rotations around the z axis of this space.

Such a result may possibly be achieved by making the fundamental wave equation invariant against all rotations in iso-space, but assuming a commutation rule which is only invariant for rotations around the z axis of this space. This procedure may possibly lead to the intended result, since the electromagnetic interactions are connected with the constant κ in the commutator, which does not appear in the wave equation. If the sign of κ is coupled with the z direction of iso-space, this direction would have been introduced into the electromagnetic interactions, but the influence on the mass values might be comparatively small.

The program indicated has not yet been followed in detail. Just as an example of how it might possibly be carried out we quote the two equations[6]:

$$L=\psi^+\gamma\mu\frac{\partial\psi}{\partial x_\mu}+\chi^+\gamma\mu\frac{\partial}{\partial x_\mu}\chi+l^2(\psi^+\chi)(\chi^+\psi) \qquad (29)$$

[19] H. Kita, Progr. Theoret. Phys. 15, 83 (1956).

[20] A. Pais, Phys. Rev. 86, 663 (1952).

[21] M. Gell-Mann, Phys. Rev. 92, 833 (1953).

[22] T. Nakano, K. Nishijima, Progr. Theoret. Phys. 10, 581 (1953).

[23] M. Goldhaber, Phys. Rev. 92, 1279 (1953) and 101, 433 (1956).

and

$$\tfrac{1}{2}S^{\psi}(p)\approx\frac{p_\mu\gamma_\mu+i\kappa\tau_3}{p^2+\kappa^2}-\frac{p_\mu\gamma_\mu+i\kappa\tau_3}{p^2}+\frac{p_\mu\gamma_\mu\kappa^2}{(p^2)^2}. \quad (30)$$

L is the Lagrangian for the wave equation and τ_3 is the z component of isobaric spin. The brackets in (21) are spinors in iso-space, but their product is an iso-scalar. The quantitative consequences of (29) and (30) have not yet been worked out. But it is easy to see what the qualitative consequences of these two equations would be with respect to the selection rules (compare Goldhaber[23]).

Besides the usual conservation laws of energy, momentum, and parity one has in (29) and (30) invariance for the transformations $\psi\rightarrow\psi e^{i\alpha}$ and $\chi\rightarrow\chi e^{i\beta}$. One of these invariances may be interpreted as conservation of the baryonic number (which is 1 for nucleons and hyperons, -1 for their antiparticles, and 0 for mesons and leptons). The rotational symmetry of (29) and (30) around the z axis of iso-space provides conservation of the z component of the isobaric spin. There will be no conservation of the total isobaric spin, since τ_3 appears in (30). But the transitions involving changes of the total isobaric spin may be somewhat less frequent than the others. Finally one has the conservation of electromagnetic charge which is connected with the introduction of the states of Hilbert space II and the conditions (19).

These selection rules together provide the existence of the "strangeness-quantum number" and its conservation. It is not necessary to invoke any new symmetry for the interpretation of this quantum number, as had been suggested by Racah[24] and Espagnat and Prentki.[25]

Therefore two rather simple equations such as (29) and (30) do actually account for all the known selection rules between baryons and mesons. Before we discuss the role of the leptons in this scheme it may be useful to form a general idea of the possible form of the weak interactions.

(b) Weak Interactions

The most characteristic feature of weak interaction between baryons and mesons seems to be the fact that it changes the z component of the isobaric spin by 1/2 without changing any of the other quantum numbers. Therefore one might think of interactions of the type

$$\text{const}\,[(\psi^+\chi)(\psi^+\psi)+\text{conj.}]. \quad (31)$$

Such expressions in the Lagrangian are scalars in ordinary space but spinors in iso-space. The constants would therefore play the role of a spinor in iso-space. This situation would lead to an interesting consequence. If one performs a rotation in iso-space by 360°, the expressions (31) change their sign, since a spinor changes its sign under this transformation. Therefore one should have written (31) with an indefinite sign $\pm$, because one cannot distinguish between the two signs. If this is true, however, one could also allow instead of (31) interactions of the different type

$$\pm\text{const}[(\psi^+\chi)(\psi^+\gamma_5\psi)+\text{conj.}]. \quad (32)$$

Such an expression would not violate the invariance of the Lagrangian for reflections in ordinary space. Still, if one combines the expressions (31) and (32), one would not expect the conservation of parity in weak interactions. Actually one of the two expressions could produce radioactive decay of the τ meson into three π mesons, while the other could cause the decay into two π mesons. Therefore this discussion favors the idea of Yang and Lee[26] that parity is actually not conserved in weak interaction. For interactions of spinor type in isobaric space [as in (31) and (32)] the invariance of the Lagrangian under reflections in ordinary space does not guarantee the conservation of parity.

(c) Role of the Leptons in the Scheme

While it is comparatively easy to combine all qualitative features of baryons and mesons in two simple equations like (29) and (30), it is difficult to see what place could be occupied by the leptons in such a scheme. The leptons are connected with the other particles both by electromagnetic interaction and by the weak interaction. Even if the weak interaction is neglected the mass spectrum of leptons should not be changed; but in this case all transitions from π mesons into leptons, β decay, etc., should be stopped. Therefore one should have a selection rule which forbids all transitions from the heavier particles into leptons except pair creation. So far no reason for such a selection rule can be given. It may be necessary to introduce a third wave function into the fundamental equation in order to account for the leptons or it may be that a closer investigation of the states in Hilbert space II would lead to a basis for such a selection rule without new wave functions. For the time being no solution for the problem of the leptons can be given.

[24] G. Racah, Nuclear Phys. 1, 301 (1956).
[25] B. d'Espagnat and J. Prentki, Nuclear Phys. 1, 33 (1956).
[26] T. D. Lee and C. N. Yang, Phys. Rev. **102**, 290 (1956).

Dielectric Spectroscopy of Ferromagnetic Semiconductors*

P. A. MILES, W. B. WESTPHAL, AND A. VON HIPPEL

Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

TABLE OF CONTENTS

* Sponsored by the Office of Naval Research, the Army Signal Corps, the Air Force, and the Army Ordnance under contracts Nonr-1841(10) and AF 33(616)-2191.

INTRODUCTION

MODERN technology depends to a large extent on ferromagnetic materials for distribution, control, and utilization of power, for high-speed computers, and electronic devices of all kinds. A new perspective for high-frequency circuitry was opened up by the development of the ferrites or ferrospinels, materials with the general formula $M^{++}Fe_2O_4$, at the Philips Research Laboratories.[1] Lately, ferromagnetic semiconductors of hexagonal,[2] as well as cubic, perovskite,[3] and garnet[4] structures have made their appearance. Like numerous other research groups, the Laboratory for Insulation Research has been active in the study of these materials over the last few years. In particular it concerns itself with their response to electric and magnetic excitations over a very broad frequency range, extending, with gaps, from dc to the ultraviolet. This review aims at a unified appraisal of the information obtained by such a broad-band approach. In the absence of an accepted name we call this effort "dielectric spectroscopy."

The stirring sequence of spectral lines, shown by individual atoms and molecules and described by the Rydberg-Ritz combination principle, led to an understanding of the quantum states of electronic excitation, vibration and rotation. These resonance spectra of individual particles extend from the x-ray into the microwave region. By condensing gases to liquids and solids, the frequency range of the electrical engineer from microwaves to dc fills up with absorption spectra, and drastic changes occur in the optical spectral range. The characteristic response of the individual particles to the electromagnetic field becomes submerged in effects of mutual interaction.

In consequence, the original sharpness of the resonance response of atoms and molecules with its revealing detail is supplanted in liquids and solids by broad resonance and relaxation spectra with much more scrambled information. Only the nuclear and paramagnetic resonance spectra may still transmit fine and hyperfine structure intelligence, because the magnetic spins are loosely coupled to their surroundings.

[1] J. L. Snoek, *New Developments in Ferromagnetic Materials* (Elsevier Publishing Company, New York and Amsterdam, 1947).
[2] Went, Rathenau, Gorter, and van Oosterhout, Philips Tech. Rev. 13, 194 (1952).
[3] G. H. Jonker and J. H. van Santen, Physica 16, 337 (1950).
[4] R. Pauthenet, Compt. rend. 242, 1859 (1956).

The spectroscopist has always considered the optical range from infrared to ultraviolet as his domain and has added in recent years microwave and magnetic spectroscopy as far as they are concerned with resonance lines and therefore amenable to his old language. The electrical engineer, by necessity, has measured the dispersion and absorption of materials from dc to microwaves, but essentially from the empirical standpoint of knowing these characteristics for applications. Only the relaxation spectra caused by the orientation of permanent electrical dipole moments, first interpreted by Debye, have become an important and widely used source of molecular information, especially for chemical problems.

Obviously, for a true understanding of liquids and solids the total range from dc to x-rays should be considered as a unified stage on which the interplay between electromagnetic fields and matter is enacted.

1. THE LANGUAGE OF DIELECTRIC SPECTROSCOPY[5]

The name "dielectric," as used here, refers to any material when viewed from the standpoint of electric or magnetic response. In their reaction to sinusoidal electric and magnetic fields, dielectrics can be characterized by their *complex permittivity* and *complex permeability*,

$$\epsilon^* = \epsilon' - j\epsilon'' \qquad \mu^* = \mu' - j\mu''. \tag{1.1}$$

These parameters, the real permittivity (dielectric constant) ϵ' and permeability μ', describing the storage of electric and magnetic energy and the loss factors ϵ'' and μ'' its dissipation, are actually measured in reference to vacuum as a *relative* complex permittivity and permeability

$$\frac{\epsilon^*}{\epsilon_0} \equiv \kappa^* = \kappa' - j\kappa'' \qquad \frac{\mu^*}{\mu_0} \equiv \kappa_m^* = \kappa_m' - j\kappa_m''. \tag{1.2}$$

In the rationalized mks system, used here, the dielectric constant and permittivity of free space are

$$\epsilon_0 = 10^7/4\pi c^2 \simeq 8.854\times10^{-12} \quad [\text{farad/m}] \qquad \mu_0 = 4\pi\times10^{-7} \simeq 1.257\times10^{-6} \quad [\text{henry/m}] \tag{1.3}$$

in which, when precision values are needed, c is the experimental value of the velocity of light in vacuum expressed in meters per second. For most practical purposes one may set $c=3\times10^8$ m/sec which is sometimes called the standardized velocity of light. From the product and ratio of these we get the velocity of light c and the impedance of free space Z

$$c \equiv (\epsilon_0\mu_0)^{-\frac{1}{2}} \simeq 3\times10^8 \quad [\text{m/sec}] \qquad Z \equiv (\mu_0/\epsilon_0)^{\frac{1}{2}} \simeq 120\pi = 376.6 \quad [\text{ohm}]. \tag{1.4}$$

Permittivity and permeability, as defined, refer to isotropic, linear dielectrics; that is, an electric field **E** creates a polarization **P** and a magnetic field **H** a magnetization **M** parallel and proportional to such field. The electric flux density **D** and the magnetic flux density **B** are the arithmetic sum of field lines and dipole chains,

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} = \epsilon^*\mathbf{E} \qquad \mathbf{B} = \mu_0\mathbf{H} + \mu_0\mathbf{M} = \mu^*\mathbf{H}. \tag{1.5}$$

The complex nature of the permittivity and permeability indicates a temporal phase shift between the acting field, **E** or **H**, and the dipole moment per unit volume, **P** or **M**, created by such field. The product of angular frequency and electric-loss factor represents a *dielectric conductivity*

$$\sigma = \omega\epsilon'' [\text{ohm}^{-1}\text{m}^{-1}], \tag{1.6}$$

which may correspond to a true conductivity created by migrating charge carriers or to any other dissipative action of the polarization.

Alternative sets of parameters, frequently used in science and engineering, can easily be expressed in terms ϵ^* and μ^*. The propagation of electromagnetic waves is described by a complex propagation factor

$$\gamma \equiv \alpha + j\beta = j\omega(\epsilon^*\mu^*)^{\frac{1}{2}}, \tag{1.7}$$

with α designating the attenuation factor and β the phase factor of the wave. The intrinsic impedance

$$Z_i \equiv \frac{E}{H} = \left(\frac{\mu^*}{\epsilon^*}\right)^{\frac{1}{2}} \tag{1.8}$$

of the dielectric represents the ratio of the coupled electric and magnetic field components as to amplitude and temporal phase for an infinitely extended material. Wavelength λ and phase velocity v in the dielectric are given as

$$\lambda = \frac{2\pi}{\beta} \qquad v = \nu\lambda = \frac{\omega}{\beta}. \tag{1.9}$$

The ratio of wavelength or phase velocity in vacuum to that in the dielectric material is the index of refraction

$$n \equiv \frac{\lambda_0}{\lambda} = \frac{c}{v} = \frac{\lambda_0}{2\pi}\beta, \tag{1.10}$$

[5] For a systematic development of the electric and magnetic concepts see A. von Hippel, *Dielectrics and Waves* (John Wiley and Sons, Inc., New York, 1954, and Technology Press, Massachusetts Institute of Technology).

and the attenuation per radian, the index of absorption,

$$k \equiv \frac{\alpha\lambda}{2\pi} = \frac{\alpha}{\beta}. \tag{1.11}$$

The indices of refraction and absorption together define the complex index of refraction

$$n^* \equiv n(1 - jk), \tag{1.12}$$

which the physicist uses for the characterization of materials in place of the propagation factor of the engineer,

$$\gamma = j\frac{2\pi}{\lambda_0}n^*. \tag{1.13}$$

The relations between flux densities and fields, measured by permittivity and permeability [see (1.5)], contain also the contribution of the field *in vacuo*. One refers to the dipole response of the material alone by introducing the electric and magnetic susceptibilities χ and χ_m:

$$\begin{aligned} \mathbf{P} &= (\epsilon' - \epsilon_0)\mathbf{E} \equiv \chi\epsilon_0\mathbf{E} \\ \chi &= \frac{\mathbf{P}}{\epsilon_0\mathbf{E}} = \kappa' - 1 \\ \mathbf{M} &= \left(\frac{\mu' - \mu_0}{\mu_0}\right)\mathbf{H} \equiv \chi_m\mathbf{H} \\ \chi_m &= \frac{\mathbf{M}}{\mathbf{H}} = \kappa_m' - 1. \end{aligned} \tag{1.14}$$

In anisotropic, linear dielectrics, **P** and **M** are not, in general, parallel to the generating fields; hence each component of **D** or **B** is a linear function of the three space components of **E** or **H** as

$$\begin{aligned} D_1 &= \epsilon_{11}E_1 + \epsilon_{12}E_2 + \epsilon_{13}E_3, \\ D_2 &= \epsilon_{21}E_1 + \epsilon_{22}E_2 + \epsilon_{23}E_3, \\ D_3 &= \epsilon_{31}E_1 + \epsilon_{32}E_2 + \epsilon_{33}E_3, \end{aligned} \tag{1.15}$$

and

$$\begin{aligned} B_1 &= \mu_{11}H_1 + \mu_{12}H_2 + \mu_{13}H_3, \\ B_2 &= \mu_{21}H_1 + \mu_{22}H_2 + \mu_{23}H_3, \\ B_3 &= \mu_{31}H_1 + \mu_{32}H_2 + \mu_{33}H_3. \end{aligned} \tag{1.16}$$

Because of the purely displacive nature of the polarization process the permittivity tensor is symmetric with $\epsilon_{ij} = \epsilon_{ji}$. The same is true of the permeability tensor at frequencies well below the inverse of the relaxation time associated with the return to equilibrium of a disturbed magnetic system. At higher frequencies, however, $\mu_{ij} = -\mu_{ji}$ for $i \neq j$. The skew symmetry of these elements reflects the gyroscopic nature of the magnetization process (Sec. 6). A possible maximum of six independent coefficients of permittivity and of permeability therefore exists. How many are not zero and different from each other, depends on the macroscopic symmetry of the system. Frequently more useful for the molecular analysis are the corresponding susceptibility matrices

$$\chi = \begin{Vmatrix} \chi_{11} & \chi_{12} & \chi_{13} \\ \chi_{21} & \chi_{22} & \chi_{23} \\ \chi_{31} & \chi_{32} & \chi_{33} \end{Vmatrix}. \tag{1.17}$$

Notwithstanding the tensor form of ϵ^* and μ^* in anisotropic crystals, the behavior of a polycrystalline aggregate of such materials is described by the simple quantities $\epsilon^* = \epsilon' - j\epsilon''$, $\mu^* = \mu' - j\mu''$, provided that the microstructure of the material is much smaller than the specimen size and the wavelength of the exciting field, and provided that the material is in a depolarized or demagnetized state.

For ferroelectrics and ferromagnetics the situation becomes more involved because the moments line up spontaneously, forming electric or magnetic axes. Changes in position cause, in general, a crystal deformation, hence a coupling exists to the elastic coefficients of the crystal structure (piezoelectricity and piezomagnetism†). Furthermore, the response to external fields is nonlinear and depends on the prehistory of the sample because domain arrays are formed to minimize the field energy.

Ferroelectrics and ferromagnetics are thus characterized by polarization and magnetization curves depending on field strength, and the permittivity and permeability coefficients here used designate certain slopes of these characteristics (Fig. 1.1). In evaluating

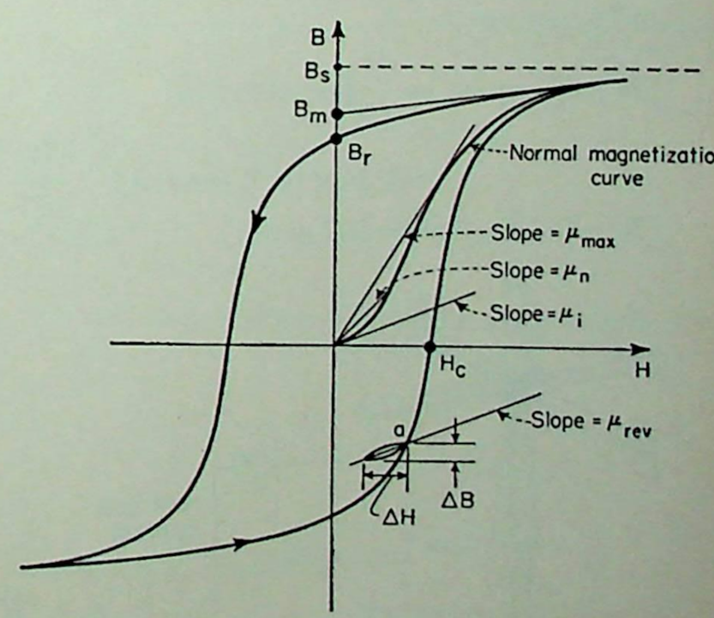


FIG. 1.1. Parameters characterizing the quasi-static magnetization of ferromagnetics.

† The term "piezomagnetism" is becoming rightfully accepted in place of "magnetostriction," because in ferromagnetics the effect is analogous to the piezoelectricity of ferroelectrics. The terms "electrostriction" and "magnetostriction" should be reserved to the mechanical deformation caused by induced moments. These latter processes have no inverse, i.e., mechanical deformation does not in this case produce or alter moments as a first-order effect.

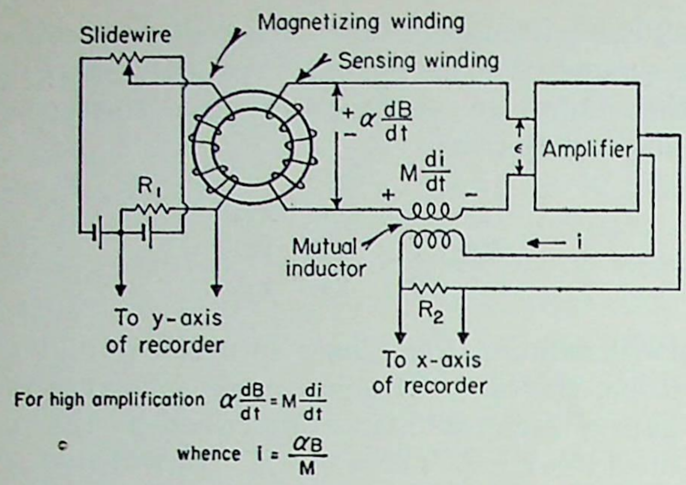


FIG. 2.1. Recording fluxmeter for quasi-static magnetic measurements.

"normal" curves and hysteresis loops, corrections are required to transform from the externally applied field strength **E** or **H** to the internally applied field strength $\mathbf{E}_i$ and $\mathbf{H}_i$. The shape of the sample enters because the free poles at the sample boundaries produce depolarizing and demagnetizing fields. For ellipsoids of revolution the internal field is homogeneous and derived from the applied field as

$$\epsilon_0\mathbf{E}_i = \epsilon_0\mathbf{E} - w\mathbf{P}$$
$$\mathbf{H}_i = \mathbf{H} - w\mathbf{M}, \tag{1.18}$$

where w is a shape factor varying from zero for a long needle parallel to the field, to one for a disk perpendicular to the field.

As mentioned at the beginning of this section, we use the rationalized mks system. The conversion factors of the magnetic field strength and flux density to the oersted and gauss are

$$[\mathbf{H}] = [\text{amp/m}] = \frac{4\pi}{10^3}[\text{oersted}]$$
$$\simeq 1.26\times 10^{-2}[\text{oersted}] \tag{1.19}$$
$$[\mathbf{B}] = [\text{weber/m}^2] = 10^4[\text{gauss}].$$

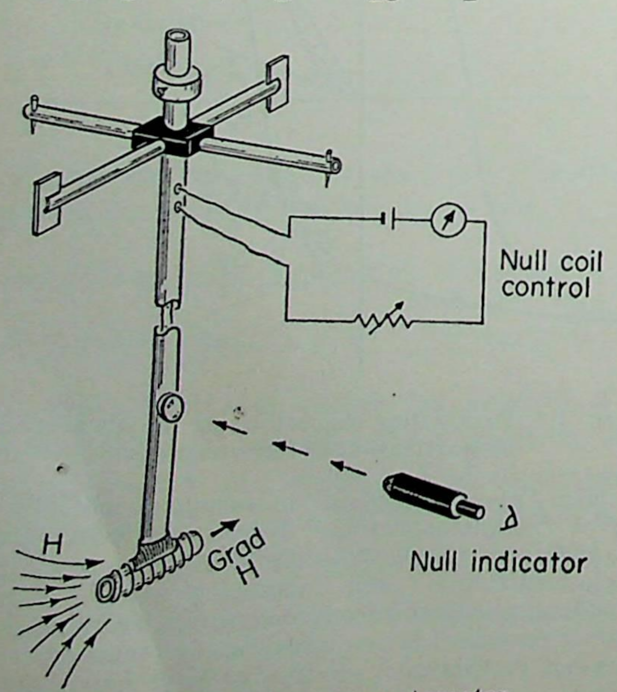


FIG. 2.2. Pendulum magnetometer.

2. MEASUREMENT TECHNIQUES

(a) Quasi-Static Magnetic Measurements

Up to ±2000 amp/m (25 oersteds), our normal magnetization curves and hysteresis loops are plotted by a quasi-static fluxmeter (Fig. 2.1), developed after a design by Cioffi.[6] For fields up to 400 000 amp/m, the normal magnetization characteristics can be determined with a null-coil pendulum magnetometer[7] (Fig. 2.2). Finally, a vibrating-coil magnetometer has been developed here recently[8] (Fig. 2.3) for rapid measurements of magnetization and anisotropy characteristics, useful over a wide temperature and pressure range. With this instrument the dipole field of the sample can be measured at a distance, hence outside of and unhindered by thermostats or pressure bombs.

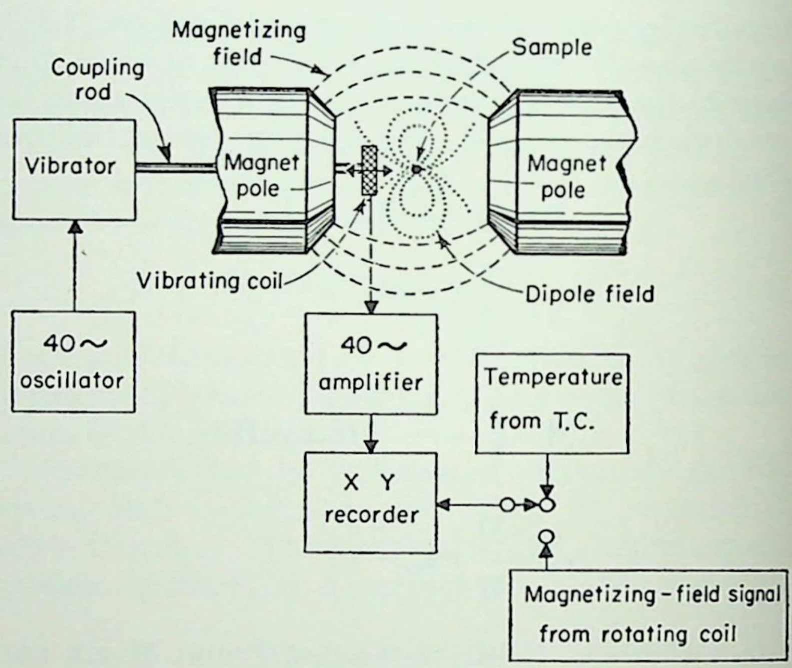


FIG. 2.3. Vibrating-coil magnetometer.

(b) Measurements of ε* by Four-Terminal Methods

To avoid boundary-layer effects, semiconductors should be measured, in general, by a four-terminal arrangement. For dc the voltage drop across the potential electrodes is compared with that across a 10 000-ohm standard resistor (Fig. 2.4). Ac measurements up to 10^4 cps are made either with a differential voltmeter circuit (Fig. 2.5) or (to avoid the nonlinearity of the electronic components) in a potentiometer circuit with the coupling transformer to the detector tuned to resonance to provide a very high input impedance (Fig. 2.6).

(c) Two-Terminal Measurements of ε* and μ* at Low Frequencies (10^2 to 10^5 cps)

In most cases the permittivity is obtained by balancing the sample capacitor in one of the bridge circuits

[6] P. P. Cioffi, Rev. Sci. Instr. **21**, 624 (1950).
[7] C. A. Domenicali, Rev. Sci. Instr. **21**, 327 (1950).
[8] D. O. Smith, Rev. Sci. Instr. **27**, 261 (1956).

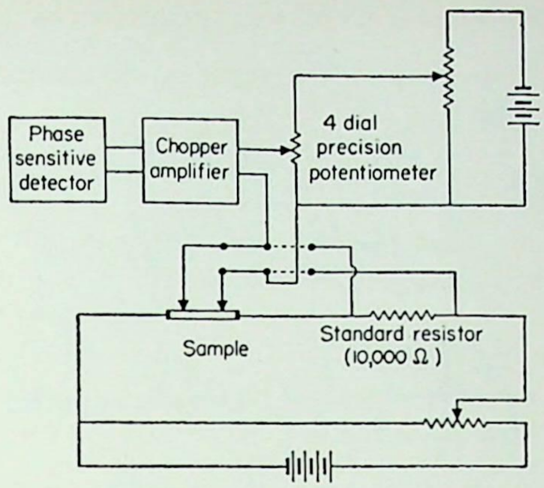


FIG. 2.4. Equipment for four-terminal dc conductivity measurements.

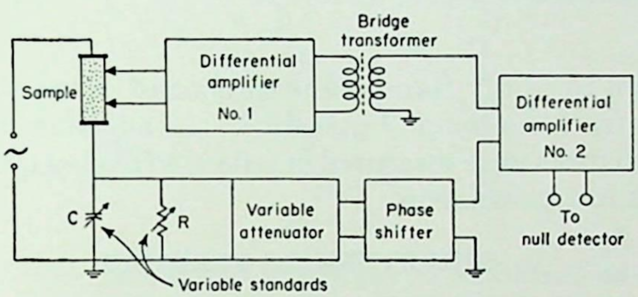


FIG. 2.5. Differential voltmeter circuit for four-terminal ac measurements up to 10^4 cps.

shown in Fig. 2.7. For permeability determinations toroid samples with windings are used and measured with a capacitance in series in the same type of bridge circuit (Fig. 2.8).

(d) ε^* and μ^* in the Medium-Frequency Range (10^5 to 10^8 cps)

Measurements of the permittivity on lumped-capacitor samples have been extended by the use of a susceptance variation circuit; dielectric constant and loss are obtained by the resonance position and half-width of the resonance curve (Fig. 2.9). For toroids in permeability determinations this technique is inferior because of the distributed capacitance of the windings. Hence the complex permeability is obtained on coaxial samples. For the range from 10^5 to 10^7 cps, such a

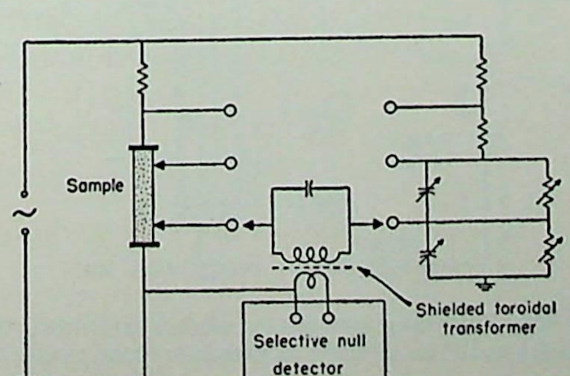


FIG. 2.6. Potentiometer circuit for four-terminal ac measurements.

sample is placed on a movable shorting plunger in a four-arm coaxial bridge (Fig. 2.10). Between 10^7 and 10^8 cps the coaxial sample can be measured at the

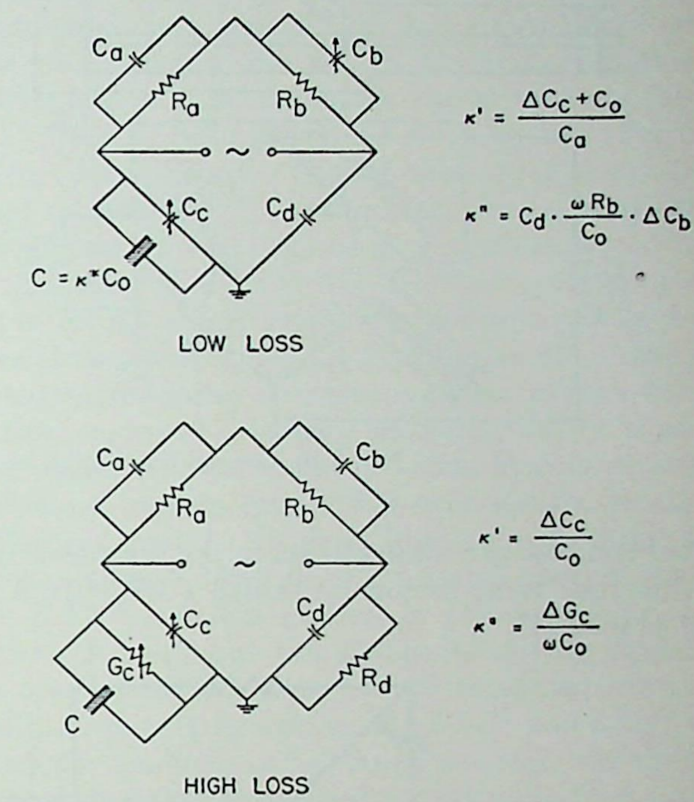


FIG. 2.7. Bridge circuits for electrical measurements from 10^2 to 10^5 cps.

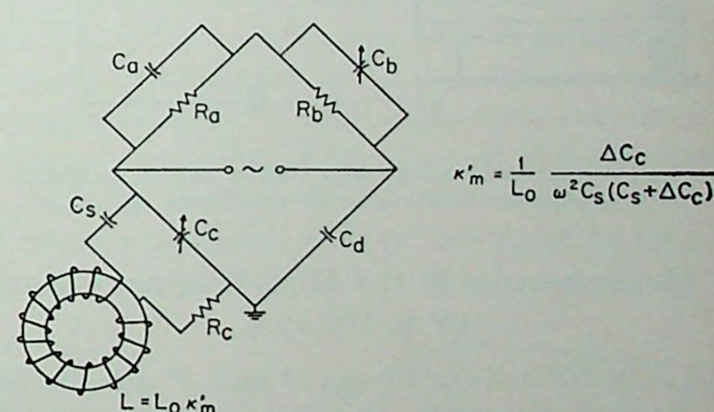


FIG. 2.8. Bridge circuit for magnetic measurements from 10^2 to 10^5 cps.

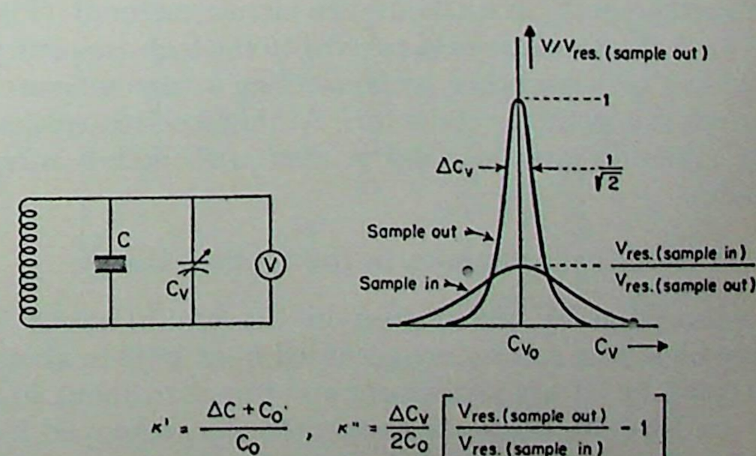
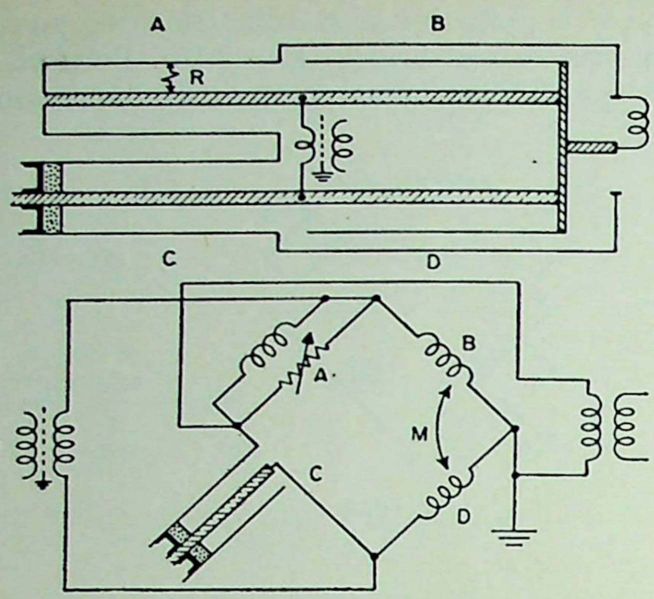


FIG. 2.9. Resonance circuit for electric measurements from 10^5 to 10^8 cps.



$$K'_m = \frac{\Delta \ell + d}{d} - \frac{\frac{\Delta \ell_R}{d}\left(\frac{Z_C}{R}\cdot\frac{2\pi\Delta \ell_R}{\lambda}\right)^2}{1+\left(\frac{Z_C}{R}\cdot\frac{2\pi\Delta \ell_R}{\lambda}\right)^2}$$

$$K''_m = \frac{\lambda}{2\pi d}\cdot\frac{\frac{Z_C}{R}\left(\frac{2\pi\Delta \ell_R}{\lambda}\right)^2}{\left(1+\frac{Z_C}{R}\cdot\frac{2\pi\Delta \ell_R}{\lambda}\right)^2}$$ Line loss neglected

FIG. 2.10. Coaxial line bridge for magnetic measurements from 10^5 to 10^7 cps.

shorted end of a fixed line section in the sample arm of a commercial radio-frequency bridge (G.R. 916-A or 1601) (Fig. 2.11).

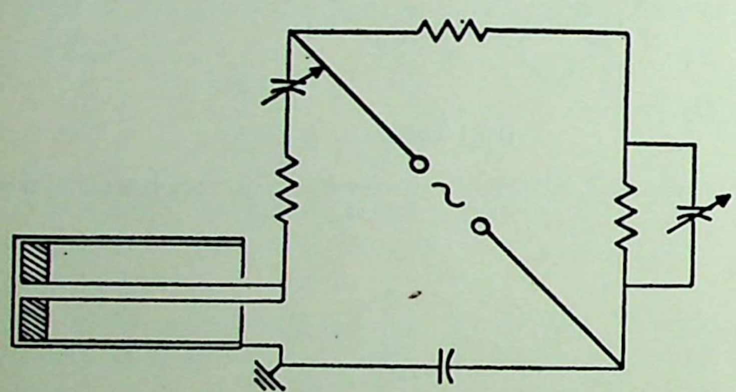


FIG. 2.11. Impedance bridge with coaxial line for magnetic measurements from 10^7 to 10^8 cps.

(e) Measurements in the High-Frequency Range (10^8 to 10^{10} cps)

Between 10^8 and 3×10^9 cps ϵ^* and μ^* are obtained in a coaxial line with standing-wave detector, by locating a thin sample in two successive measurements in a region of high magnetic and high electric field strength (short-circuit and open-circuit method) (Fig. 2.12). A dc magnetic field parallel to the high-frequency field can be superposed by connecting a current source across the center conductor. At higher frequencies, the standing-wave method is used with hollow wave guides.

(f) Measurements in the Optical Range

Absorption by the ferrites in the optical range is very high. Our spectra are recorded from 2000 to about 10 000 A by a Cary instrument and from 1 to about 40 μ by an IR-3 Beckman spectrophotometer, equipped for the far infrared with KRS-5 optics. For a survey study, ceramic or crystal samples were pulverized to particles of the order of 0.1 to 1 μ, mixed with KBr powder and pressed to nearly transparent specimens[9] in the form of disks (*ca* 1 cm diam, 2 mm thick). In addition, single-crystal disks were measured in reflection and evaporated layers in transmission.

3. STRUCTURE OF FERRITES

The ferrites considered here have the spinel structure, a closest-packed cubic array of O^{2-} spheres with compensating cations in interstitial positions. Adjacent to each oxygen ion are 14 interstices: 6 located in the cube-edge directions, surrounded by six oxygen ions (octahedral sites), 8 placed in the space-diagonal directions, surrounded by four oxygen ions (tetrahedral sites) (Fig. 3.1).

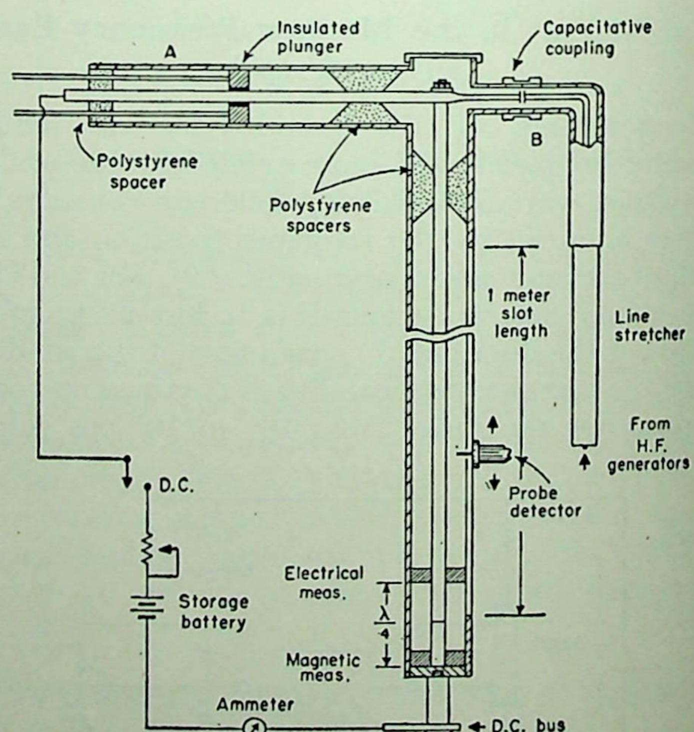


FIG. 2.12. Standing-wave method for electric and magnetic measurements from 10^8 to 10^{10} cps in presence of dc magnetic bias.

[9] M. M. Stimson and M. J. O'Donnell, J. Am. Chem. Soc. 74, 1805 (1952); U. Schiedt and H. Reinwein, Z. Naturforsch. 7B, 270 (1952).

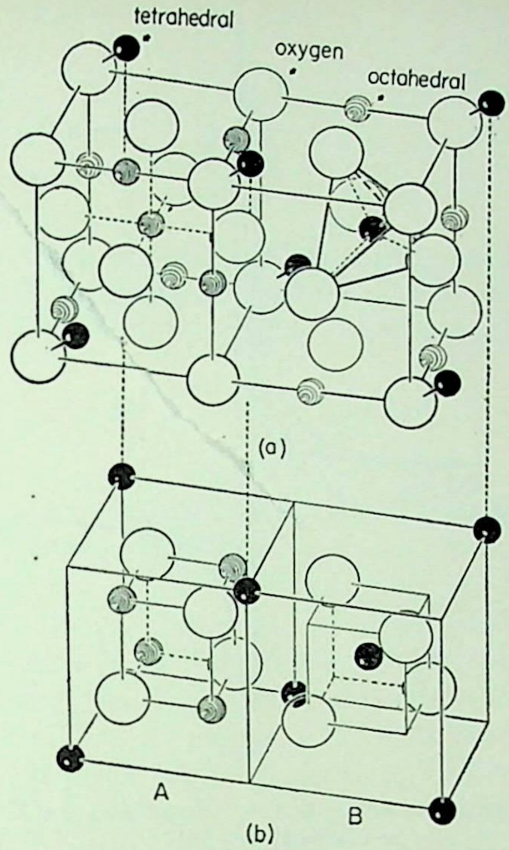


Fig. 3.1. Two views of spinel structure: (a) octahedral and tetrahedral sites; (b) repeating sections of unit cell.

The prototype of these ferromagnetic spinels is magnetite (Fe_3O_4). Here each anion is balanced by $\frac{3}{4}$ of a cation; hence, placing around each oxygen "a" cations in the octahedral and "b" cations in the tetrahedral positions, we have to fulfill the condition

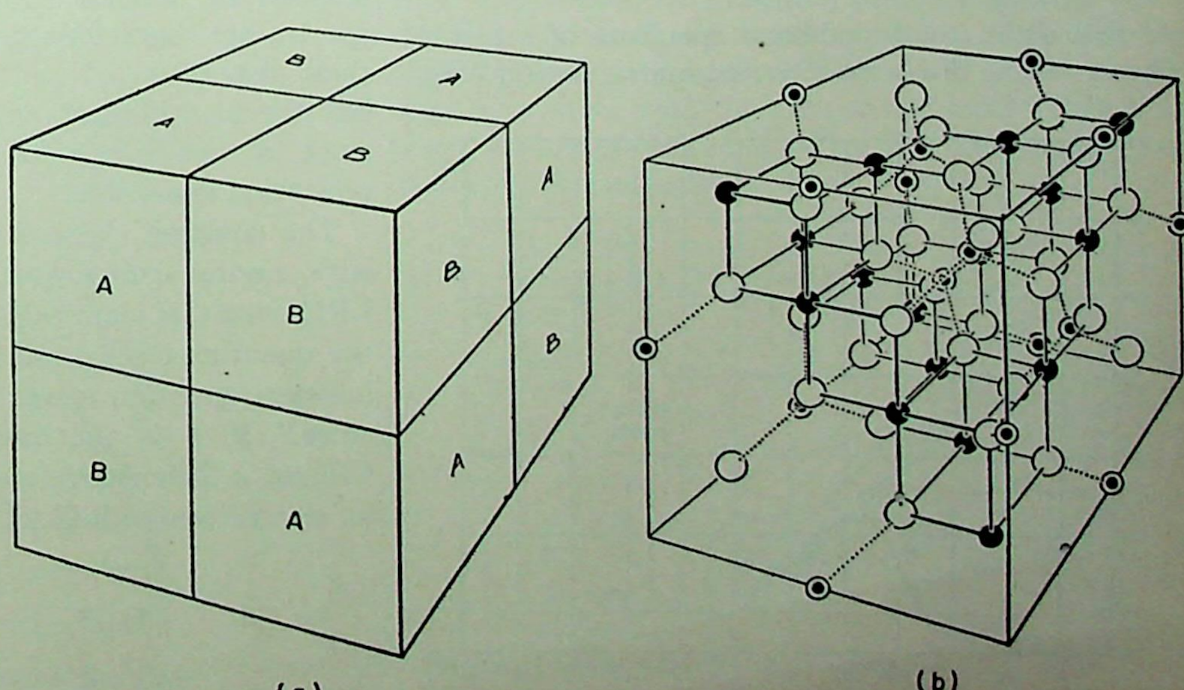
$$\frac{a}{6}+\frac{b}{4}=\frac{3}{4}; \tag{3.1}$$

that is, $a=3$, $b=1$ or $a=0$, $b=3$. The cations repel each other and therefore should be placed at maximum possible separation. If, in addition, we require that a repetition of the oxygen-cation configuration in space should produce a symmetrical and stable lattice structure, the observed unit cell of magnetite results with 32 oxygens and 24 cations in the arrangement $a=3$, $b=1$. The unit cell comprises 8 Fe_3O_4 groups (Fig. 3.2).

Thus far we have assumed that three iron ions of equal valence ($Fe^{(8/3)+}$) compensate the negative charge of four oxygen ions. Actually, a chemist would write for magnetite $Fe_3O_4=Fe^{2+}O^{2-}Fe_2{}^{3+}O_3{}^{2-}$, that is, postulate 8 ferrous and 16 ferric ions per unit cell. In structures of the spinel type, the distribution of cations over the permissible sites may lie anywhere between the two extreme cases of 8 divalent ions on tetrahedral sites and 16 trivalent ions on octahedral sites (normal spinel) or 8 trivalent ions on tetrahedral sites and the remaining 8 divalent and 8 trivalent ions on octahedral sites (inverse spinel). The physicist finds[10] that, in magnetite, 8 Fe^{3+} ions are located in tetrahedral (A) sites (inverse spinel), but the sixteen remaining cations in the octahedral (B) sites can be separated only artifically at room temperature into 8 Fe^{3+} and 8 Fe^{2+} ions. The high conductivity of magnetite suggests that the electron exchange between the ferrous and ferric ions at identical lattice sites is so rapid ($\sigma \simeq 10^2$ ohm^{-1} cm^{-1}), that an average charge of $+2.5$ should be assigned to these iron atoms in octahedral positions.

Each iron ion is the carrier of a magnetic moment; hence the two different cation sites form two magnetic sublattices A and B. Néel[11] successfully explained the spontaneous magnetizations observed in simple ferrites by assuming that the predominant coupling between the magnetic moments of the cations is that between the

Fig. 3.2. Spinel unit cell: (a) stacking pattern of repeat units; (b) lattice.



[10] E. J. W. Verwey, Nature 144, 327 (1939); E. J. W. Verwey and P. W. Haayman, Physica 8, 979 (1941).
[11] L. Néel, Ann. phys. [12] 3, 137 (1948).

tetrahedral and octahedral sublattices and that this $A \leftrightarrow B$ coupling is antiferromagnetic. Thus the saturation magnetization of magnetite at absolute zero should correspond to that of the ferrous ions only:

	B sites	*A* sites	
Filled cation sites per unit cell:	Octahedral (16)	Tetrahedral (8)	Resultant moment
Magnetic moments:	8 $Fe^{++}\uparrow$ 8 $Fe^{+++}\uparrow$	8 $Fe^{+++}\downarrow$	8 $Fe^{++}\uparrow$.

The form of the magnetization *vs* temperature curve, on the other hand, depends essentially on the magnitude and nature of the couplings within each sublattice. In general these are found to be antiferromagnetic, although magnetite appears to have a weak ferromagnetic coupling between the ions of sublattice *A*.

Near 119°K, one observes an electronic order-disorder transition in magnetite. The electron exchange between Fe^{2+} and Fe^{3+} has obviously slowed down so much that the lattice can respond to the whereabouts of the electrons by distortion. Thus an ordered $Fe^{2+}Fe^{3+}$ structure can be frozen in at the *B* sites. In Néel's sense, we should now distinguish between one *A* and two *B* magnetic sublattices.

By substituting other cations for Fe^{2+} and/or Fe^{3+}, a great variety of materials of ferrite-type can be made in which the resulting magnetic properties depend in a complicated manner on the spin interactions in and between such sublattices. A satisfactory physical analysis of the ferrites requires accurate knowledge of the cation distribution, of the electron exchange between multivalent ions, of the origin and coupling of the magnetic moments, and an understanding of their interaction with the crystal lattice. In this connection it must be considered that a purely ionic description of the structure may not be justified.

4. PRINCIPLES OF DIELECTRIC ANALYSIS

At first sight, the broad-band spectrum of a ferrite evokes a feeling of complexity and inarticulation (Fig. 4.1). A number of effects are superposed and guiding principles are needed to reduce the integrated characteristics into meaningful components.

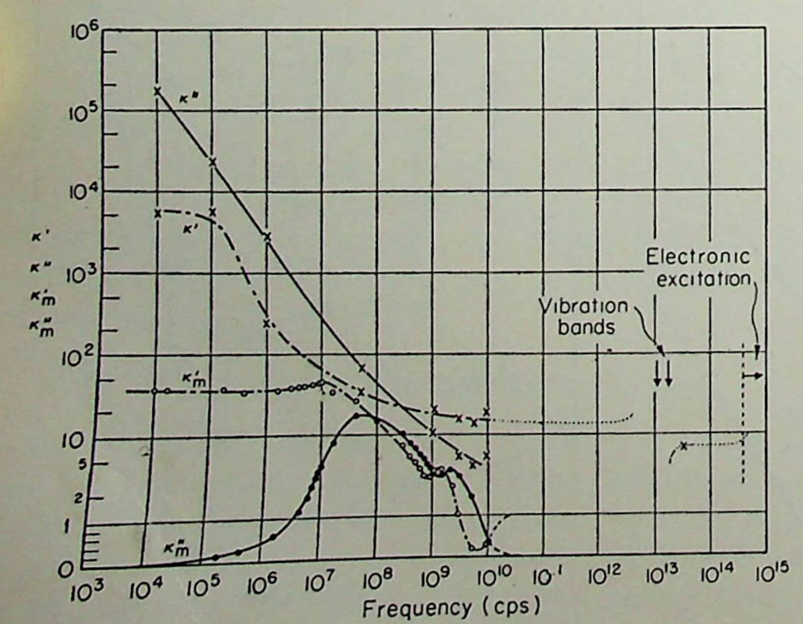


FIG. 4.1. Broad-band dielectric spectra of nickel ferrite (dispersion and absorption).

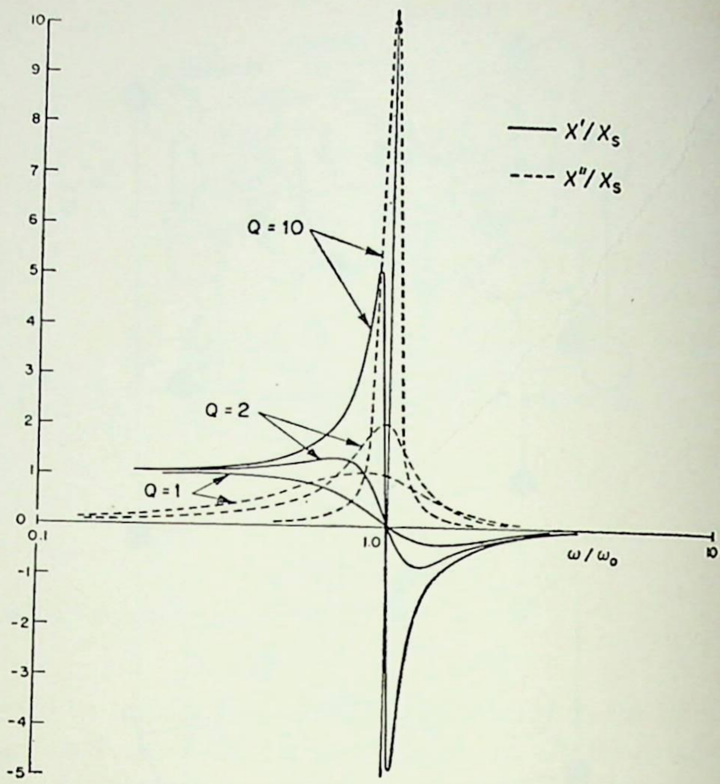


FIG. 4.2. Normalized susceptibility (dispersion and absorption) of classical resonator.

The response to electromagnetic fields is under consideration. Hence it is logical that we inquire about electric processes excited by electric fields, magnetic processes excited by magnetic fields, and interaction phenomena between the two. Furthermore, electromagnetic fields couple to matter essentially by dipole terms; hence charge carriers, induced moments, and permanent moments will make their appearance and respond in resonance and relaxation spectra. These spectra are characterized by the frequency response of their dispersion (κ') and absorption (κ'') components, but the identification as to type may be ambiguous, since a continuous transition from resonance to relaxation is possible.

The common classical resonator, a mechanical one with viscous damping or, electrically, a simple series LRC circuit, is identical in its frequency response with the quantum-mechanical resonator with its transitions between quantum states of isolated atoms and molecules.[12] If N is the number of resonators per unit volume, e their charge and m their mass, one observes an electric susceptibility [see (1.14)]

$$\chi^* = \kappa^* - 1 = \frac{Ne^2/\epsilon_0 m}{\omega_0^2 - \omega^2 + j\omega 2\alpha}, \tag{4.1}$$

or, with respect to the static susceptibility χ_s, a nor-

[12] See reference 5, pp. 161 ff.

malized susceptibility (Fig. 4.2)

$$\frac{\chi^*}{\chi_s}=\frac{1}{1-\left(\frac{\omega}{\omega_0}\right)^2+j\frac{\omega}{\omega_0{}^2\tau_l}}. \qquad (4.2)$$

For low damping, the spectral line is symmetrical about the resonance frequency ω_0 with a width $\Delta\omega_h$ at half-power points given by the attenuation factor 2α

$$\Delta\omega_h=2\alpha=\frac{1}{\tau_l}. \qquad (4.3)$$

τ_l is the *lifetime* of the excited state. In classical terms, the quality factor **Q** of the resonator, defined as the ratio of energy stored to that dissipated per half-cycle, is equal to $\omega_0/\Delta\omega_h$. Transitions between quantized magnetic states in isolated systems, described by the magnetic susceptibility, follow the same form.

Elimination of the acceleration term in the resonance equation leads to the simplest prototype of a relaxation spectrum (Fig. 4.3)

$$\frac{\chi}{\chi_s}=\frac{1}{1+j\omega\tau}, \qquad (4.4)$$

where the *relaxation time* τ corresponds, for example, to Debye's ordering time of permanent dipoles in a viscous medium.[13] We can transform from (4.2) to

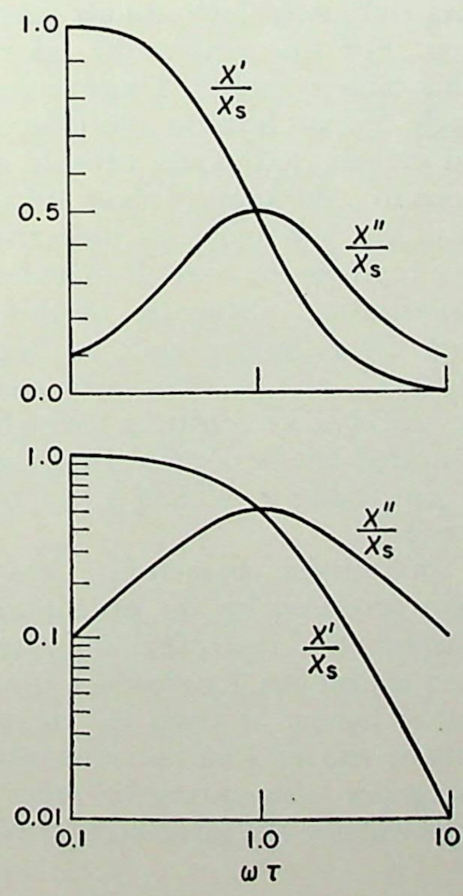


FIG. 4.3. Dielectric relaxation spectrum (dispersion and absorption).

[13] P. Debye, *Polar Molecules* (Chemical Catalog Company, New York, 1929), pp. 83 ff.

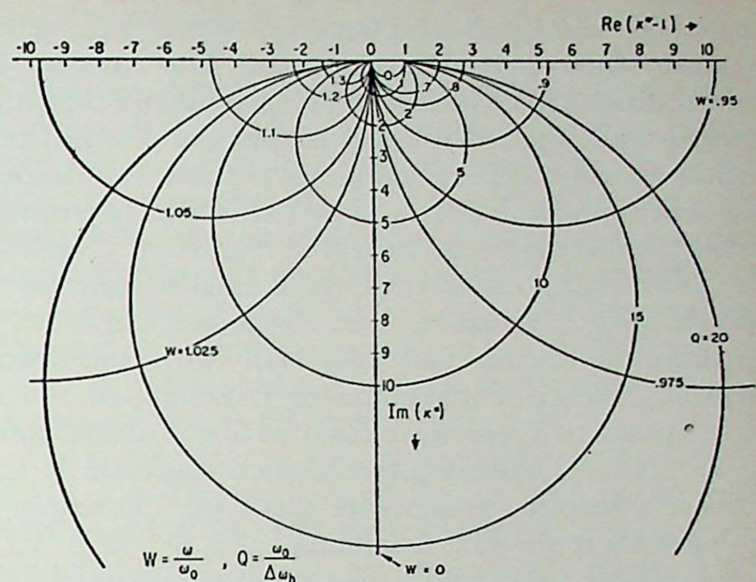


FIG. 4.4. Vector plot of the complex susceptibility (κ^*-1), illustrating continuous transition from resonance to relaxation spectrum.

(4.4) by defining formally

$$\tau=\frac{1}{\tau_l\omega_0{}^2}, \qquad (4.5)$$

shifting the resonance to very high frequencies in comparison to the frequency of observation $(\omega\ll\omega_0)$, and increasing the damping α, that is, reducing the lifetime τ_l until $\tau_l\ll(1/\omega_0)$. The physical meaning of this procedure is that we wipe out the discrete quantum states of the individual particles by interaction with their surroundings, and observe the adjustment of the particles in these surroundings. A typical example is the pressure broadening of spectral lines when a disrupting collision interferes before a vibration or rotation cycle can be completed. A plot of the vector χ^* (Fig. 4.4) shows the transition from a resonator (with the phase of its dipole moment relative to that of the inducing field varying from 0° to 180°) to the simple relaxing system with its circular arc locus (Cole-Cole plot[14]) and phase shift from 0° to 90°. The circular loci of Fig. 4.4 have the form

$$\left[\chi'-\frac{1}{2(1-w^2)}\right]^2+\chi''^2=\frac{1}{4(1-w^2)^2},$$

$$w=\frac{\omega}{\omega_0}=\text{constant}; \qquad (4.6)$$

and

$$\left[\chi'-\frac{1-\sqrt{w}}{1-w^2}\right]^2+\left(\chi''-\frac{Q}{2}\right)^2\simeq\frac{Q^2}{4} \quad \text{for} \quad Q>2.$$

Electron transitions and vibrations of individual particles find their counterpart in the electron transi-

[14] K. S. Cole and R. H. Cole, J. Chem. Phys. 9, 341 (1941).

tions and vibrations of coupled systems and hence maintain their resonance character. The rotation of electric dipoles in condensed systems leads to relaxation spectra, while the much looser coupling of electron and nuclear spins to their surroundings permits well-defined magnetic resonance spectra to persist. The experimental distinction between resonance and relaxation (Fig. 4.4) may be difficult, but two criteria, if fulfilled, unequivocally identify resonance. The dispersion curve of a relaxation spectrum can only fall with increasing frequency, hence a maximum or minimum of the κ' curve attests to a resonance contribution. Furthermore, the κ'' curve of the simple relaxation spectrum is the steepest absorption possible for this type. Plotted on a log-log scale it shows a maximum slope ± 1 (Fig. 4.3). Hence, a loss peak with a slope steeper than 1 is an unmistakable sign of resonance.

5. ELECTRIC SPECTRA

(a) Electronic Transitions

The absorption spectra of bound electrons in solids consist normally of a broad sequence of bands in the "eigenabsorption" region, correlated to specific transitions. They are preceded towards longer wavelengths by weaker absorptions arising from impurities and distorted lattice areas. Light absorption, especially at the tail of the intrinsic band, may lead to photoelectric conduction and to the buildup of new absorptions towards the infrared. Electronic conduction, to which the quantum-mechanical conduction band concept applies, will produce a continuum which underlies the other absorptions. Figure 5.1 illustrates these types of spectra for the case of impure silicon.

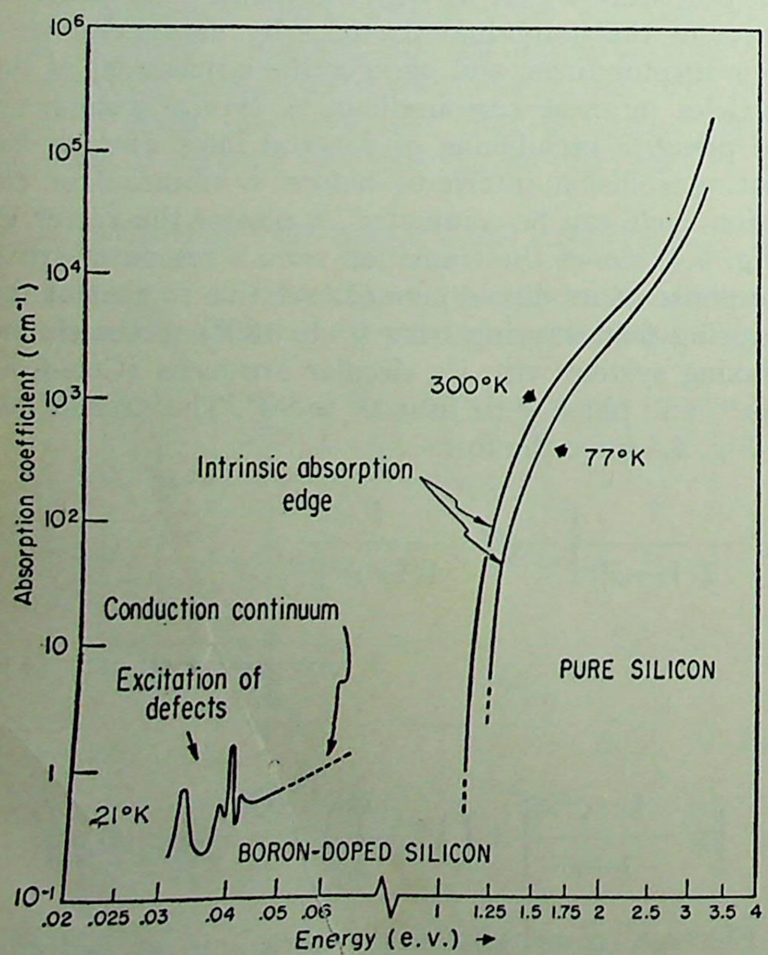


FIG. 5.1. Electronic absorption spectrum of silicon [W. C. Dash and R. Newman, Phys. Rev. 99, 1151 (1955); R. Newman, *ibid.* 103, 103 (1956)].

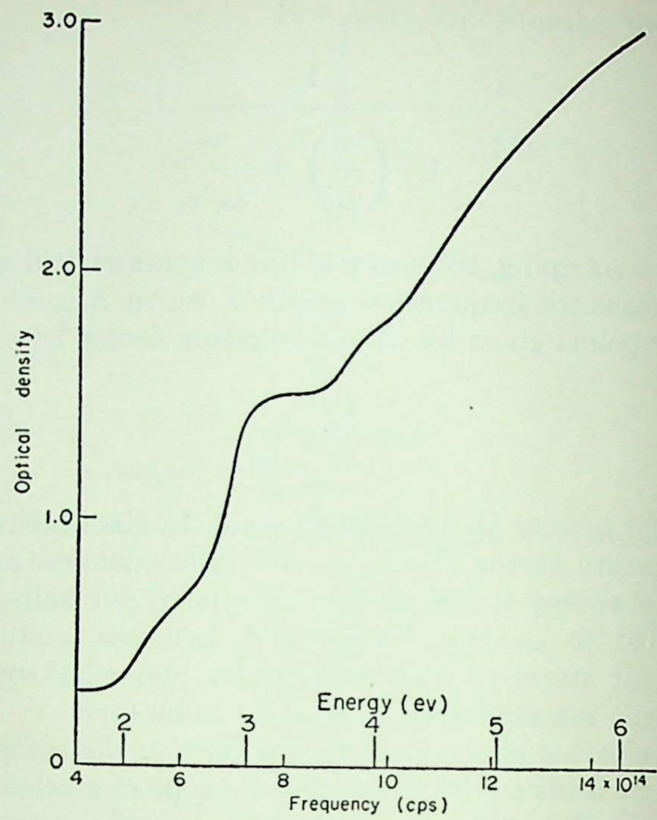


FIG. 5.2. Transmission of magnetite film in the optical region.

The ferrites considered here absorb strongly in the optical region. For this reason the information we possess at this time is limited. Transmission measurements on single crystals have to await development of grinding and etching techniques capable of reducing crystal sections to a thickness of about 0.1 micron. For a discussion of the low-frequency dielectric constant, it is necessary to know the over-all contribution of the electronic polarization. Estimates of this value are obtained from the refractive index of Mg ferrite at 0.6 micron ($\kappa' \simeq 6.5$) and by reflectivity measurements on a single crystal of nickel ferrite on the high-frequency side of the infrared bands ($\kappa' \simeq 7$). Thus the electronic polarization contributes a value of 6 to 7 to the static dielectric constant.

An actual transmission characteristic was taken on a magnetite film produced by the evaporation of pure iron and its oxidation to Fe_3O_4 (Fig. 5.2). Its scrambled information on transitions, for example, from cation $3d$ levels to higher bands, or from O^{2-} to the various cations, points up the need for measurements on single crystals at very low temperatures in order to achieve resolution sufficient for the identification of individual lines.

(b) Infrared Vibrations

The unit cell of magnetite, as mentioned in Sec. 3 (Fig. 3.2), contains eight Fe_3O_4 groups or 56 particles,

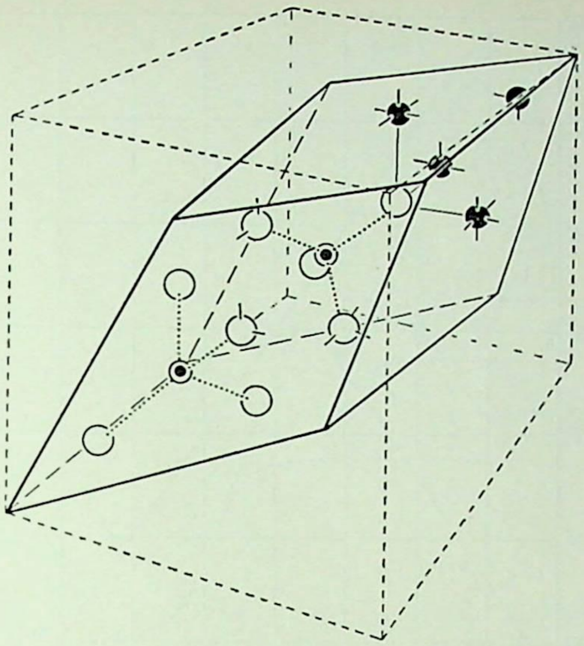


FIG. 5.3. Rhombohedral subcells of spinel lattice.

each of which contributes three degrees of freedom. Hence the unit cell has 168 normal vibrations. This inconveniently large number can be cut down by a factor of four by dividing the unit cell into four primitive rhombohedral subcells (Fig. 5.3). If we write the formula of the ferrites as MFe_2O_4, where M stands for a divalent metal ion, the rhombohedral cell ($M_2Fe_4O_8$) contains 14 ions which can be lumped into two MO_4 and one Fe_4 tetrahedral subgroups.

Of the 42 possible modes those which create a change in dipole moment $[(\partial\mathbf{u}/\partial r)\neq 0]$ are optically active. Only eight vibratory modes cause local dipole moment changes; three of these become inactivated because the over-all dipole-moment change in the primitive cell is zero; a fourth is a zero-frequency translation. Hence, four optically active modes remain and are in phase throughout the lattice, when the subcells are joined together.[15]

The mode of highest frequency (ν_1) corresponds to the vibrations of the oxygen ions along the tetrahedral bond directions, the [111] crystal directions; the next lower mode (ν_2) represents a motion of the oxygen ions in a direction almost perpendicular to the first one; the remaining two modes (ν_3, ν_4) correspond to oscillations of the metal ions in the force fields of their octahedral and tetrahedral oxygen environments (Fig. 5.4).

In absorption measurements, the real part α of the propagation constant γ is obtained. For a nonmagnetic material, γ may be expressed in terms of the dielectric constant $\kappa^*=\kappa'-j\kappa''$ as [Eq. (1.7)]

$$\gamma=\alpha+j\beta=\gamma_0\left(\frac{|\kappa^*|-\kappa'}{2}\right)^{\frac{1}{2}}+j\gamma_0\left(\frac{|\kappa^*|+\kappa'}{2}\right)^{\frac{1}{2}}. \quad (5.1)$$

However, when the material is powdered and imbedded in a loss-free dielectric, the absorption coefficient of the resultant mixture is measured. This is not simply the coefficient for the material, reduced in magnitude by the dilution, but depends on other factors. For example, pronounced scattering will occur when the particle sizes become of the order of the wavelength; the dielectric constant of the embedding medium also enters.[16] In the present work scattering becomes of importance only in the visible range. For particles very much smaller than the wavelength, the absorption is expected to follow the variations in the dielectric loss factor κ'' for the ferrite and to show peaks near the vibrational resonance frequencies.

Absorption spectra for six simple ferrites are shown in Fig. 5.4. Two main bands occur, with peaks in the vicinity of 550 and 400 cm^{-1}, which we attribute to the two highest modes; the two lower bands apparently lie beyond the range of our instrument. $ZnFe_2O_4$ is a "normal" spinel, that is, the Zn^{2+} ion occupies the tetrahedral lattice sites, while Fe_3O_4, $NiFe_2O_4$, and

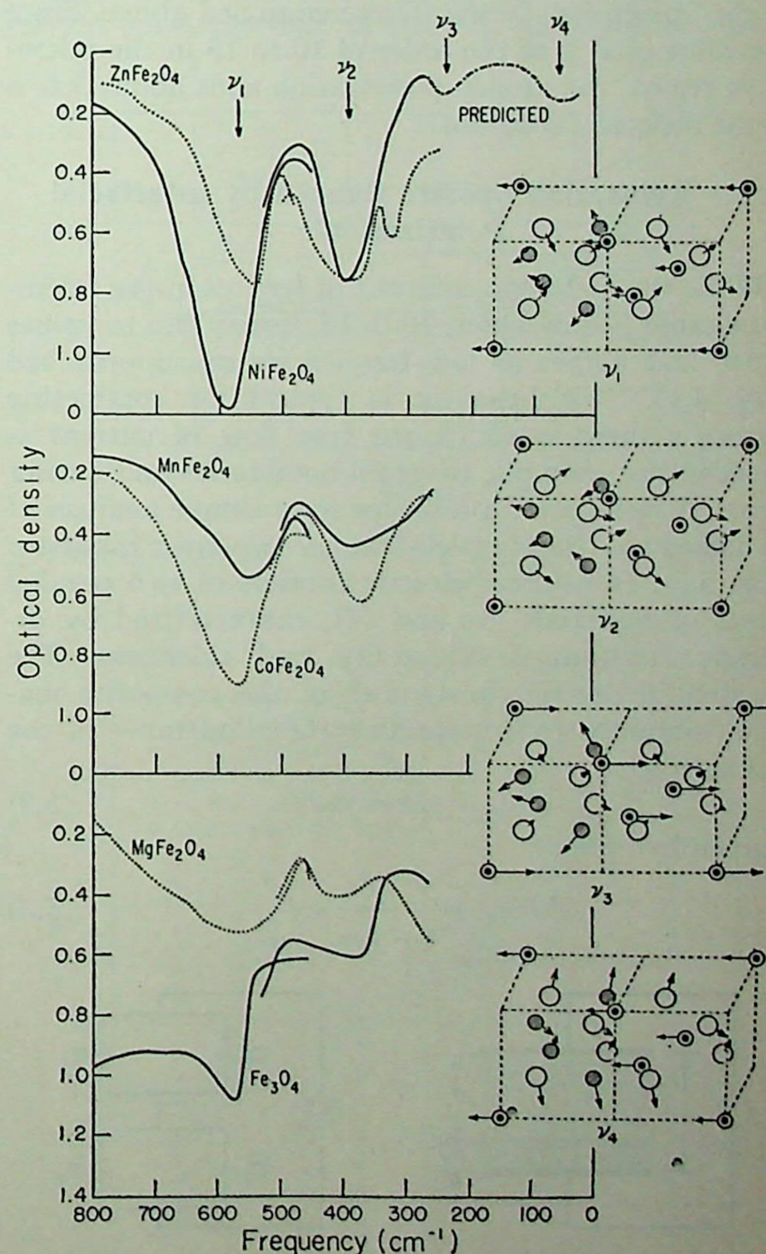


FIG. 5.4. Vibration spectra and associated active modes of ferrites (Waldron[15]).

[15] R. D. Waldron, Phys. Rev. 99, 1727 (1955).

[16] J. T. Last, Phys. Rev. 105, 1740 (1954).

$CoFe_2O_4$ are "inversed" spinels. $MgFe_2O_4$ and $MnFe_2O_4$ are intermediate cases, on which firing conditions and cooling have great influence.[17] The spectra show that, in the inversed structure, additional charge in the tetrahedral site increases the energy and restoring force associated with the bonds between the tetrahedral atom and its oxygen neighbors. This increases the frequency separation of ν_1 and ν_2 by 30 to 40 wave numbers. In addition, slight changes are expected because of the lowering of symmetry, but the large experimental band widths limit the sensitivity of this method of analysis.

An attempt was made to map the complete dispersion and absorption characteristics of nickel ferrite in the infrared, by measuring the polarization of the reflected light as a function of the angle of incidence on a single-crystal specimen. Unfortunately, the accuracy in this long wavelength range proved insufficient for reliable tracing of such curves. As a significant result, a value of κ', 7.0 ± 0.5, was obtained at the high-frequency side of the absorption ($\lambda\simeq 5\ \mu$), as mentioned above. Since the value of κ' is of the order of 10 to 15 in the microwave region, the atomic polarization adds about 3 to 8 to the dielectric constant.

(c) Relaxation Spectra Caused by Interfacial Polarization[18]

While the dielectric constant of ferrites in the microwave range is only about 10 to 15, it may rise to values of 10^4 and higher as low frequencies are approached (Fig. 4.1). This behavior is typical for conducting materials through which the free flow of current is impeded, for example, by grain boundaries and barrier layers. A well-known prototype for a simple analysis of such effects is the Maxwell-Wagner two-layer capacitor (Fig. 5.5). Here the dielectric consists of two parallel sheets of materials (1) and (2), characterized by dielectric constant, conductivity and thickness. The apparent dielectric constant κ^* of the composite material, obtained by interpreting the admittance of the capacitor as

$$Y_{\text{cap}} = j\omega C = j\omega\kappa^* C_0 \tag{5.2}$$

is given by

$$\kappa^* = \kappa_\infty' + \frac{\kappa_s' - \kappa_\infty'}{1 + j\omega\tau} - j\frac{\sigma}{\omega\epsilon_0}, \tag{5.3}$$

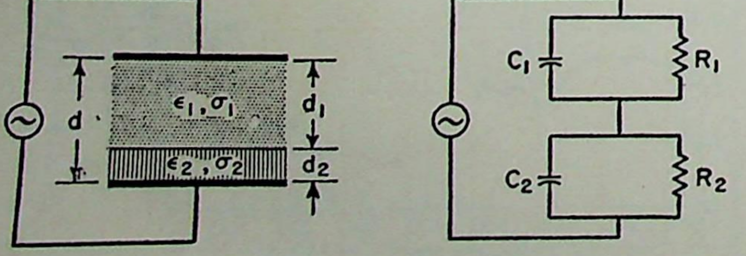


FIG. 5.5. Maxwell-Wagner two-layer capacitor and its equivalent circuit.

[17] G. Economos, J. Am. Ceram. Soc. 38, 282, 335 (1955).
[18] See reference 5, pp. 228 ff.

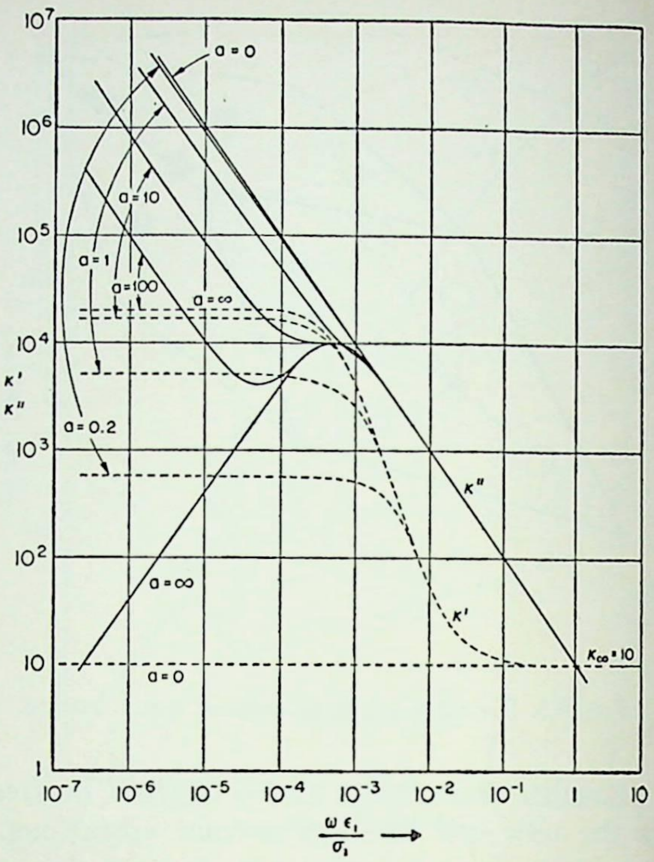


FIG. 5.6. Dielectric spectra of interfacial polarization.

where

$$\kappa_\infty' = \frac{(d_1+d_2)\epsilon_1\epsilon_2}{(d_1\epsilon_2+d_2\epsilon_1)\epsilon_0},$$

$$\kappa_s' = \frac{(d_1+d_2)}{\epsilon_0}\cdot\frac{(\epsilon_1 d_1\sigma_2^2+\epsilon_2 d_2\sigma_1^2)}{(\sigma_1 d_2+\sigma_2 d_1)^2},$$

$$\sigma = \frac{(d_1+d_2)\sigma_1\sigma_2}{(d_1\sigma_2+d_2\sigma_1)},$$

and

$$\tau = \frac{\epsilon_1 d_2+\epsilon_2 d_1}{\sigma_1 d_2+\sigma_2 d_1}.$$

Examination of (5.3) shows that the dispersion caused by the double layer is identical with that of the dipole relaxation of (4.4), while the absorption contains an additional term, $\sigma/\omega\epsilon_0$, dominant at low frequencies, due to the series resistors R_1 and R_2. In order to see the dependence of the dispersion and absorption characteristics on the relative properties of the two layers, we introduce factors of proportionality a and b, where

$$a = \frac{R_2}{R_1} = \frac{\sigma_1 d_2}{\sigma_2 d_1}, \quad b = \frac{C_2}{C_1} = \frac{\epsilon_2 d_1}{\epsilon_1 d_2}. \tag{5.4}$$

Figure 5.6 shows typical characteristics with the factor b chosen arbitrarily as 2000 and κ_∞' as 10, orders of magnitude frequently observed. The effect of increasing resistance ratios is illustrated by the parameter a.

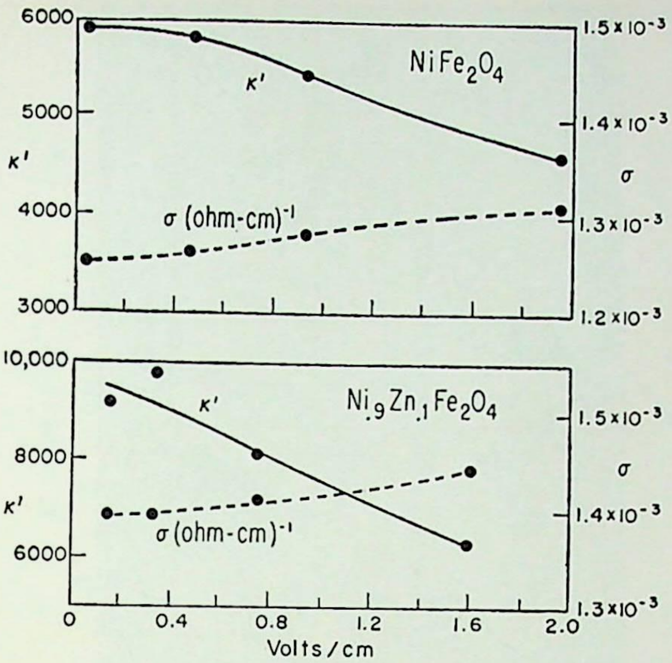


FIG. 5.7. Field-strength dependence of the interfacial polarization at 10^4 cps.

The two-layer model can be altered into an n-layer model of two media or into a dispersion of spherical particles of medium 1, distributed in a matrix of medium 2. As long as the relative amounts of media 1 and 2 remain the same, the complex permittivity stays unaltered. However, the prerequisite "spherical" is needed in order to preserve a simple relaxation spectrum with only one time constant.

Returning with this information to the electrical spectra of ferrites, we find, as did Koops,[19] that we can approximately match the observed characteristics (Figs. 7.1 and 7.2) by the equivalent circuit of Fig. 5.6 by choosing, for (5.4), parameter a between 0.1 and 1.5 and parameter b between 10^4 and 2×10^5. The ferrite proper (medium 1) is assumed to be dispersed as spherical particles in an interface (medium 2) which forms thin layers of high resistivity around the particles; hence $C_2\gg C_1$, $R_2\simeq R_1$, $\tau_2\gg\tau_1$. In consequence, $b\gg1$, $a\simeq1$, and at sufficiently high frequencies the interface medium is shortened out, and one observes the dielectric properties of the base material.

For magnesium ferrite-aluminates a comparison with the two-layer model has been made by Fairweather and Frost[20]; constants a and b were derived from 2-terminal measurements. We find that 2-terminal data can be quite unreliable in the lower frequency range and may show a field-strength dependence which varies with the type of electrode used. Four-terminal measurements eliminate these difficulties. The remaining consistent field-strength dependence (Fig. 5.7) is probably caused by the high electric stress across the thin layers of medium 2, which leads to localized field emission and breakdown.

(d) Spectra of Ionic Atmospheres

After the relaxation spectrum of the interfacial polarization has been traversed toward increasing frequency we should expect a characteristic which remains constant until the infrared absorption becomes noticeable. Instead, the dielectric conductivity begins to rise again appreciably between 10^9 and 10^{11} cycles (Fig. 5.8). A similar behavior is observed for glasses when they contain loosely bound cations (Fig. 5.9). Furthermore we find that while the microwave conductivity of a typical nickel-zinc ferrite increases slowly with temperature, the dc conductivity shows the exponential decrease expected for a normal semiconductor (Fig. 5.10).

This indicates that an additional absorption spectrum, lying in the 10^{11} to 10^{12} cps range, moves towards lower frequencies as the temperature decreases. Only the flat region of the σ characteristic is therefore descriptive of the electronic conductivity of the ferrites, while the bulk of the electric losses at micro-

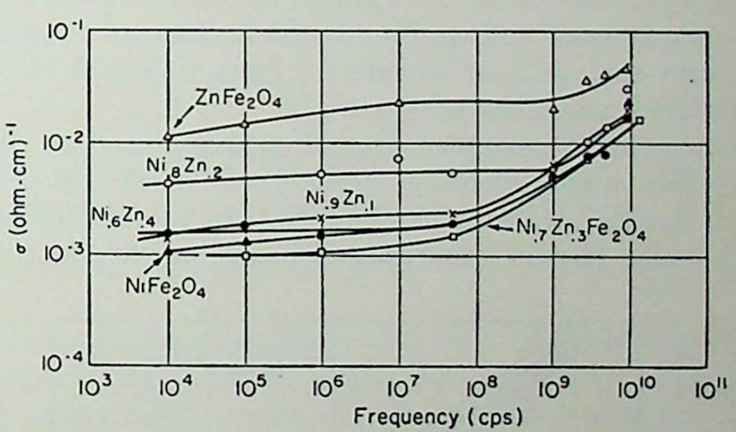


FIG. 5.8. Increase of the dielectric conductivity of ferrites towards the microwave region.

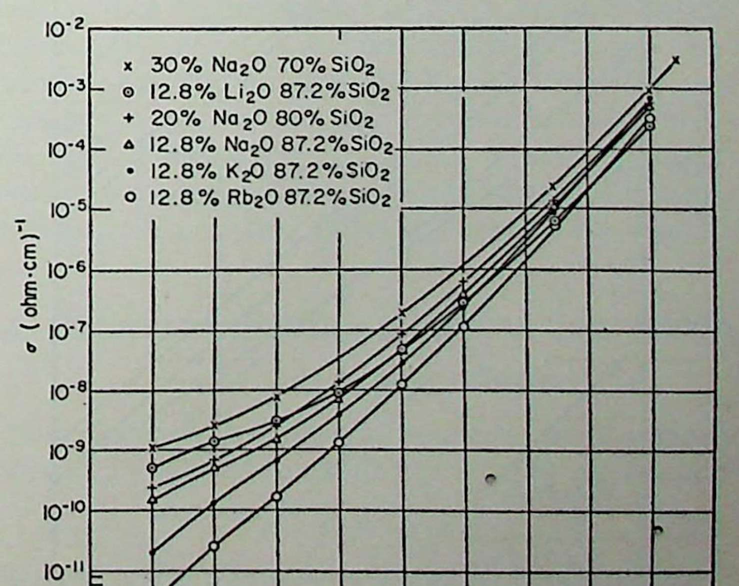


FIG. 5.9. Increase of the dielectric conductivity of glasses towards the microwave region.

---

[19] C. G. Koops, Phys. Rev. **83**, 121 (1951).
[20] A. Fairweather and E. J. Frost, Proc. Inst. Elec. Engrs. **100**, Pt. IIA, 15 (1953).

wave frequencies arises from the incoming absorption spectrum.

This absorption is possibly caused by loosely bound charge carriers moving out of phase with their compensating ionic surroundings. Such spectra would be akin to the relaxation spectra of ionic atmospheres in liquids[21] (Wien effect), but might have resonance character in solids.

(e) The Conductivity of Ferrites

Interfacial polarization effects of ferrite ceramics certify that the observed conductivity may not at all be the true conductivity of the base material. There may be a variety of such effects, depending on composition and pre-treatment. For highly resistive ferrites, for example, an exponential increase of the conductivity with temperature has been reported to be governed by the same activation energy below and above the relaxation region.[20] In this case, the transition layers appear to be not of different electric properties but rather to be represented by geometrical intergranular contact areas. On the other hand, true transition layers frequently exist and can be identified, e.g., as a highly resistive surface layer of different texture (Fig. 5.11) in a sintered manganese-zinc sample.

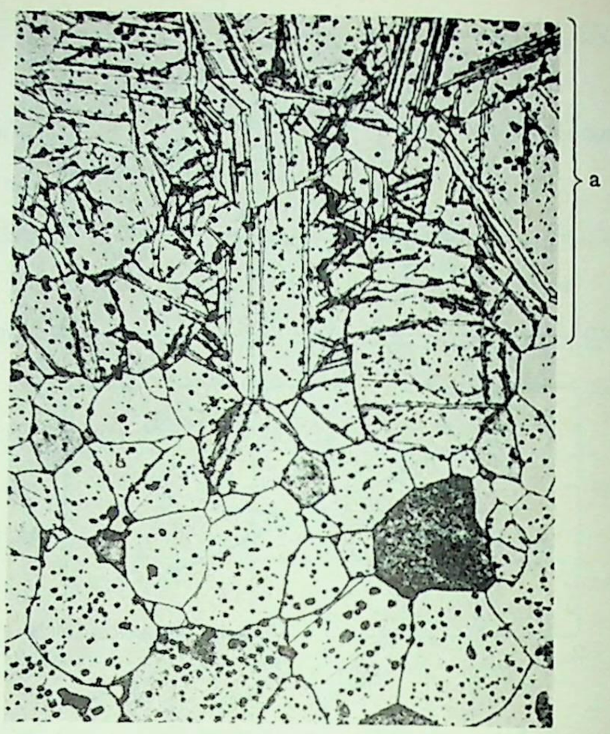


Fig. 5.11. Highly resistive surface layer (a) of manganese-zinc ferrite.

Our research samples are, in general, fired to appreciably higher density (95–99% of x-ray density) than commerical samples (75–85%). In the case of our nickel-zinc ferrites this resulted in an increase in conductivity by about three orders of magnitude, even in the microwave region. The reason is probably that at high temperatures electron transitions take place from O^{2-} to Fe^{3+}. An excess of Fe^{2+} is created and the over-all charge balance restored by oxygen loss from the material. In cooling, a porous sample can reabsorb oxygen with relative ease from the firing atmosphere, and return to a low-temperature ferric equilibrium. For dense samples, the diffusion of the oxygen is too slow and a greater excess of ferrous iron remains.

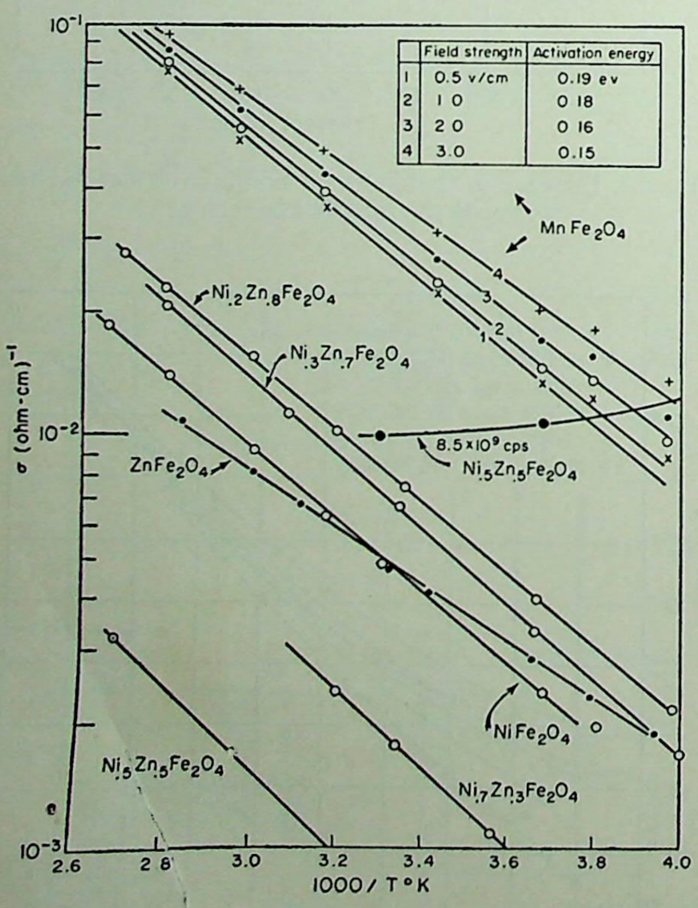


Fig. 5.10. Temperature dependence of the dc conductivity of ferrites (Epstein[45]).

This explanation is consistent with a number of other observations. Kamiyoshi[22] found that the activation energies of nickel ferrite and cobalt ferrite can be decreased considerably by quenching from high temperature. Wijn[23] showed that oxygen deficiency in a Ni-Zn ferrite can increase its conductivity from 10^{-6} to 10^{-3} (ohm cm)$^{-1}$ and lower the activation energy from 0.4 to 0.1 ev. Finally, measurements on single crystals of magnetite[24] have given us detailed information on the dominant role played by the $Fe^{2+} \rightarrow Fe^{3+}$ electron exchange in the conduction process.

The temperature dependence of the dc conductivity of a magnetite single crystal shows a very unusual characteristic (Fig. 5.12). One striking feature is an abrupt drop of the conductivity, by several orders of magnitude, when the crystal is cooled through a transition from cubic to orthorhombic symmetry near 119°K. This effect was interpreted by Verwey and co-workers[10] as an electronic order-disorder transition in the $Fe^{2+}-Fe^{3+}$ occupation of the octahedral cation

[21] P. Debye and H. Falkenhagen, Physik. Z. 29, 121 (1928); M. Wien and co-workers, Physik. Z. 37, 155 (1936).

[22] K. Kamiyoshi, Sci. Repts. Research Inst. Tôhoku Univ. 3, 716 (1951).

[23] H. P. J. Wijn, Lab. Philips Gloeilampenfabrieken, Separaat 2092, 1953.

[24] C. A. Domenicali, Phys. Rev. 78, 458 (1950); B. A. Calhoun, *ibid.* 94, 1577 (1954).

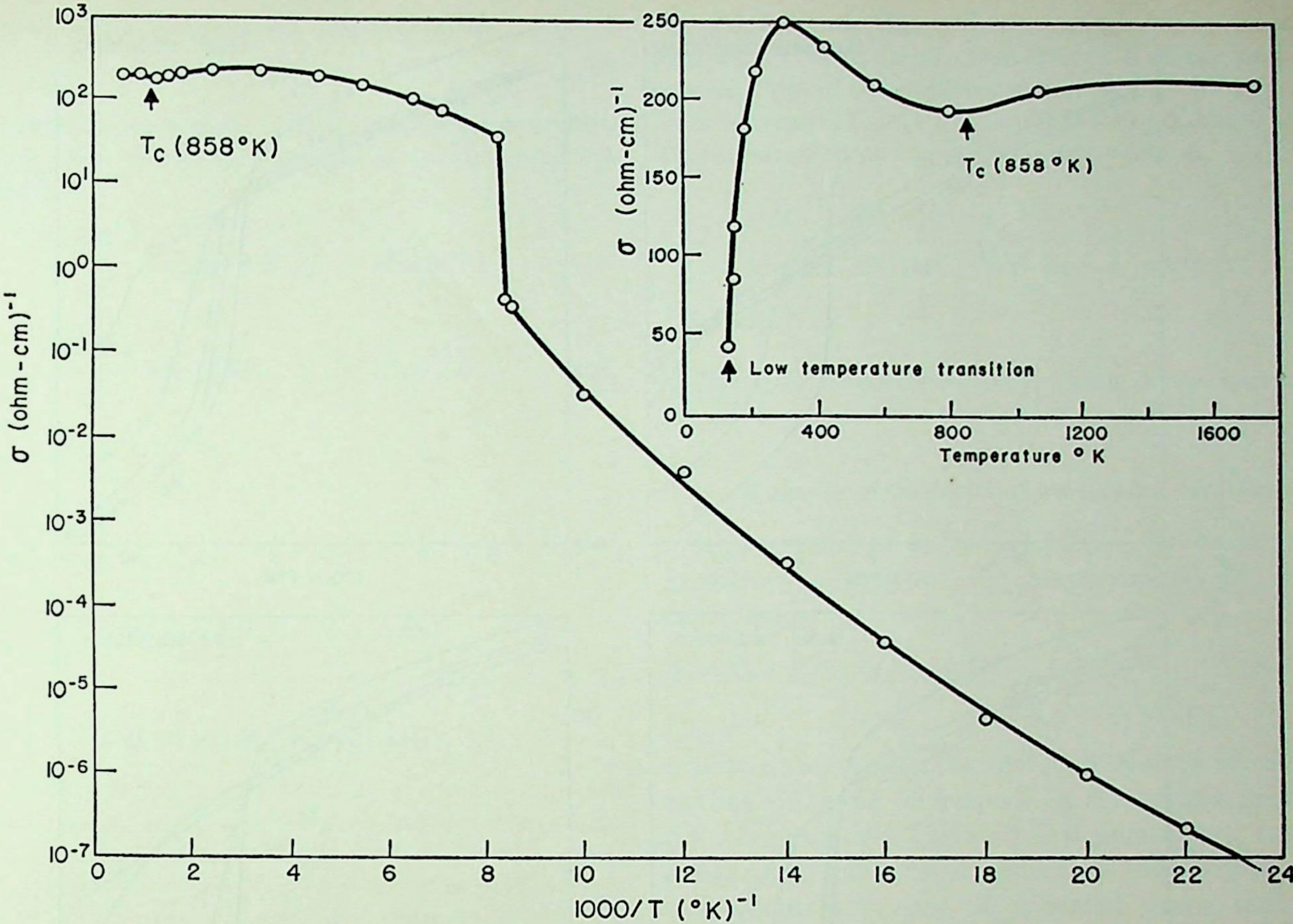


FIG. 5.12. Temperature dependence on the dc conductivity of the magnetite single crystal.

sites. Observations on a concurrent ordering of the magnetic spin system[24] support this explanation. The sharpness of the transition is rapidly destroyed by addition agents which interrupt the ferrous-ferric sequence in the octahedral sites (Fig. 5.13).

Without doubt the very high conductivity of magnetite is caused by the easy electron transfer in the octahedral sites, and the highly resistive ferrites are made by interrupting this transfer. The electron exchange in magnetite may be considered as an activation process, in which the electrons are handed along so fast above the transition that the vibrations cannot follow. At the transition, the lattice vibrations catch up and differentiate between ionic arrangements by different distortion of the structure at the Fe^{2+} and Fe^{3+} sites. Thus, above the transition only the electronic polarization enters into the activation process; below it, both the electronic and atomic polarization are involved. The ordering is not complete immediately below the transition, as conduction measurements show,[24] but improves with falling temperature; also the apparent activation energy, as measured by the slope of log conductivity *vs* $1/T$ curve, decreases slowly from 0.1 ev and reaches 0.03 ev at 40°K.

It is tempting to speculate that in magnetite single crystals the unusual case of band-type conduction in specific lattice directions [110] is realized, which changes below a critical temperature to a more classical trapping and release process. More research is needed to clarify this issue. At higher temperature, scattering by lattice vibrations flattens the slope of the conductivity characteristic and reverses it near 355°K. The new rise above the Curie point may be caused by strong onset of the higher energy electron transfer processes between O^{2-} and the cations postulated above. Such a transfer would simultaneously weaken the superexchange coupling causing ferromagnetism and be a determining factor for the Curie temperature. Electron transfer between cations on tetrahedral and octahedral sites may also become of importance in the disordered magnetic state.

6. MAGNETIC SPECTRA (GENERAL DISCUSSION)

(a) Gyroscopic Effects

The elementary magnetic moments **m** of a material are casually related to quantized angular momenta **p**′ as

$$\mathbf{m}=\gamma\mathbf{p}', \tag{6.1}$$

where γ is the gyromagnetic ratio. Viewed classically, they are gyroscopes, subjected in a magnetic field **H**, according to Newton's law, to a torque

$$\mathbf{T}=\frac{d\mathbf{p}'}{dt}=\mathbf{m}\times\mu_0\mathbf{H}, \tag{6.2}$$

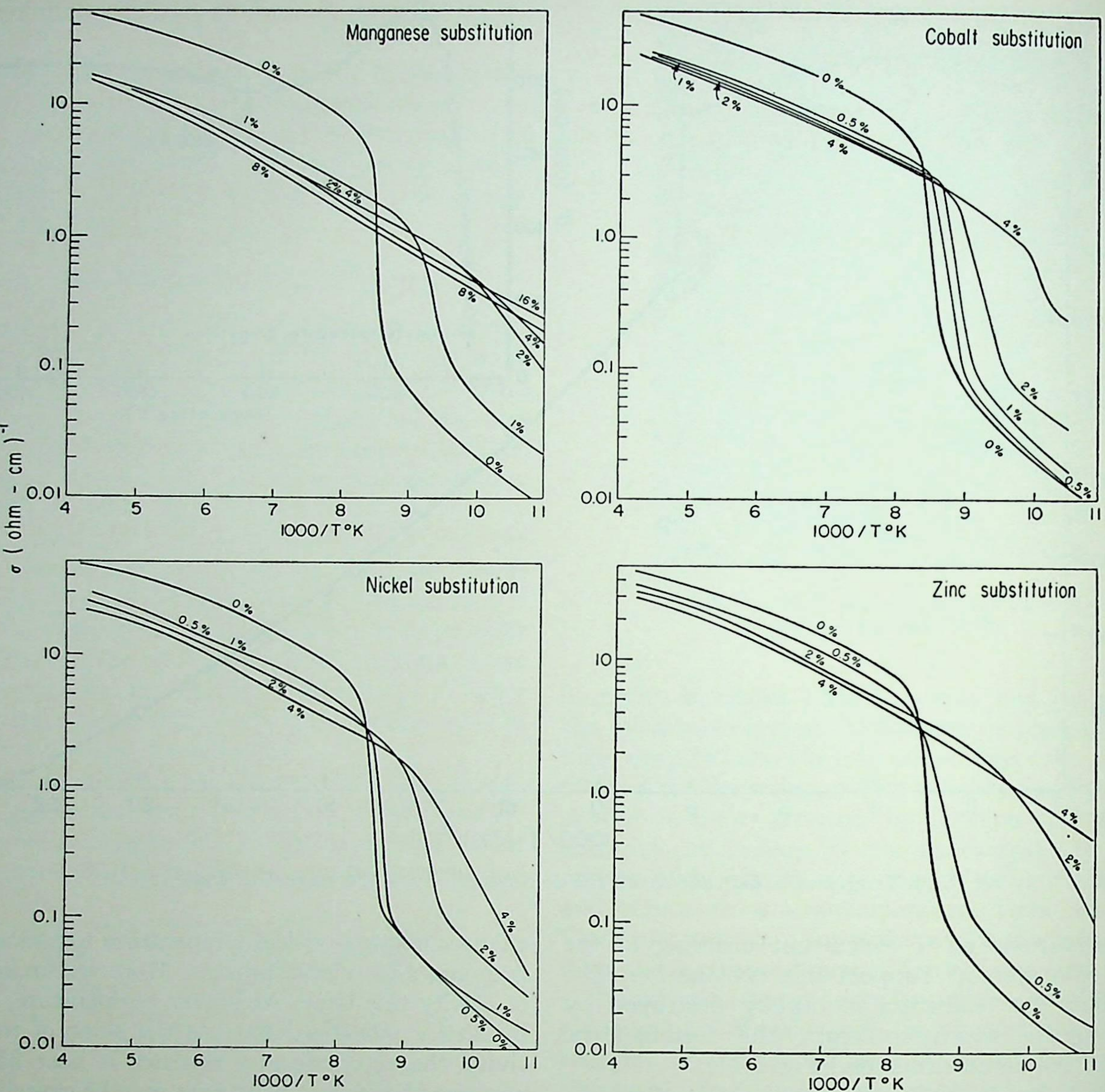


FIG. 5.13. Effect of cation substitution on the electronic order-disorder transition in magnetite (J. H. Epstein, M. S. thesis, Massachusetts Institute of Technology, 1954).

which causes a temporal change of the magnetic moment

$$\frac{d\mathbf{m}}{dt}=\gamma[\mathbf{m}\times\mu_0\mathbf{H}]. \tag{6.3}$$

The magnetization, represented by the additive action of N dipole moments per unit volume,

$$\mathbf{M}=N\mathbf{m}, \tag{6.4}$$

therefore obeys the equation of motion

$$\frac{d\mathbf{M}}{dt}=\gamma[\mathbf{M}\times\mu_0\mathbf{H}]. \tag{6.5}$$

Precession in Static Fields

If $\mathbf{H}$ represents a static field applied in the $+z$ direction,

$$|\mathbf{H}|=H_z=\text{constant}, \tag{6.6}$$

(6.5) has the solution

$$\begin{aligned} M_x &= M_0 e^{j\omega_0 t} \\ M_y &= M_0 e^{(j\omega_0 t+\pi/2)} \\ M_z &= \text{constant}=(M^2-M_0^2)^{\frac{1}{2}}, \end{aligned} \tag{6.7}$$

with $\omega_0=-\gamma\mu_0 H_z$. The two oscillating components are out of phase by 90° in space and time, hence add up to a circular rotation in the x-y plane. The magnetization vector $\mathbf{M}$ precesses around the magnetic field axis. (The sign of ω_0 indicates a counterclockwise precession for $\gamma>0$, as viewed looking along the direction of the applied field.)

Superposition of a Small Alternating Magnetic Field

If the magnetization vector $\mathbf{M}$ is initially at rest, pointing in the direction of the static field H_z, the super-

position of a small alternating magnetic field

$$\Delta \mathbf{H} = (\Delta \mathbf{H})_0 e^{j\omega t} \tag{6.8}$$

again causes a precession. Without attenuation and by including only first-order terms, the torque equation (6.5) reads

$$\begin{aligned} \frac{d(\Delta M_x)}{dt} &= \gamma\mu_0[(\Delta M_y)H_z - M_z(\Delta H_y)] \\ \frac{d(\Delta M_y)}{dt} &= \gamma\mu_0[M_z(\Delta H_x) - (\Delta M_x)H_z] \\ \frac{d(\Delta M_z)}{dt} &= 0, \end{aligned} \tag{6.9}$$

the solution of which is

$$\begin{aligned} \Delta M_x &\equiv a(\Delta H_x) - j\frac{\omega}{\omega_0}a(\Delta H_y); \\ \Delta M_z &\equiv a(\Delta H_y) + j\frac{\omega}{\omega_0}a(\Delta H_x), \end{aligned} \tag{6.10}$$

with

$$a \equiv \frac{\omega_0^2 \dfrac{M_z}{H_z}}{\omega_0^2 - \omega^2}.$$

Expressing the relation between the magnetization and field components by a susceptibility matrix for the magnetic susceptibility $\chi_m{}^*$, we obtain $\chi_{11} = \chi_{22} = a$, and $\chi_{12} = -\chi_{21} = -j(\omega/\omega_0)a$,

$$\chi_m = \begin{Vmatrix} a & -j\dfrac{\omega}{\omega_0}a & 0 \\ j\dfrac{\omega}{\omega_0}a & a & 0 \\ 0 & 0 & 0 \end{Vmatrix}. \tag{6.11}$$

An experiment designed to measure the diagonal susceptibilities χ_{nn} would therefore trace out the dispersion curve

$$\frac{\chi_{11}}{\chi_s} = \frac{\chi_{22}}{\chi_s} = \frac{\omega_0^2}{\omega_0^2 - \omega^2}, \tag{6.12}$$

a resonance curve identical with that for an undamped linear oscillator (Sec. 4).

(b) Spin Resonance and Effective Field

In Eq. (6.7) we assumed that the Larmor frequency is determined by the applied field H_z; this is true only in isotropic magnetic media without domain and crystal boundaries. More generally, the resonance frequency can be represented as proportional to an "effective" field, $\mathbf{H}_{\text{eff}}$, which contains all influences that lead to a change in the free energy $\mathcal{E}$ of the magnetic system, when the magnetization vector $\mathbf{M}$ is turned from its equilibrium direction.[25] Writing the free energy $\mathcal{E}$ as a function of the polar angles θ and ϕ,

$$\begin{aligned} \omega_0 &= -\frac{\gamma\mu_0}{M \sin\theta}\left[\frac{\partial^2\mathcal{E}}{\partial\theta^2}\cdot\frac{\partial^2\mathcal{E}}{\partial\phi^2} - \left(\frac{\partial^2\mathcal{E}}{\partial\theta\cdot\partial\phi}\right)^2\right]^{\frac{1}{2}} \\ &= -\gamma\mu_0 H_{\text{eff}}. \end{aligned} \tag{6.13}$$

The most important contributions of the material to the effective field are the following.

Magnetic Anisotropy of the Crystal Structure

In a crystal of cubic symmetry, for example, the anisotropy energy of the magnetization $\mathbf{M}$ can be expressed as

$$\mathcal{E}_a = K_0 + K_1(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2) + K_2(\alpha_1^2\alpha_2^2\alpha_3^2) + \cdots, \tag{6.14}$$

where α_1, α_2, α_3 are the direction cosines of the magnetization vector in respect to the crystal axes (Fig. 6.1). Landau and Lifshitz[26] first pointed out that this anisotropy leads to an effective field, which for a single-domain crystal of spherical shape, when the first-order term dominates,[27] amounts to

$$H_{\text{eff}} = \frac{2K_1}{\mu_0 M_s}$$

for K_1 positive (easy direction [100])

and (6.15)

$$H_{\text{eff}} = -\frac{3K_1}{2\mu_0 M_s}$$

for K_1 negative (easy direction [111]).

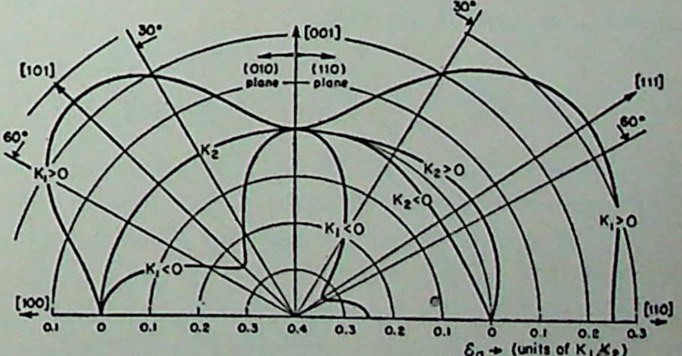


FIG. 6.1. Magnetic anisotropy energy in two planes of a cubic crystal.

[25] H. Suhl, Phys. Rev. **97**, 555 (1955).
[26] L. Landau and E. Lifshitz, Physik. Z. Sowjetunion **8**, 135 (1931).
[27] C. Kittel, Phys. Rev. **73**, 155 (1948).

Shape Factor

For ellipsoids the position of minimum energy for an isotropic ferromagnetic is the direction of the long axis, because the smallest demagnetizing field results (1.18). Hence, deflection of the magnetization vector from this direction gives rise to a torque tending to restore the minimum energy state. In this case the effective field for a single-domain ellipsoid[27] polarized in the z direction by a field H_z is given by

$$\mathbf{H}_{\text{eff}}=[H_z+(w_x-w_z)M_z]^{\frac{1}{2}}[H_z+(w_y-w_z)M_z]^{\frac{1}{2}} \equiv H_{x\text{eff}}{}^{\frac{1}{2}}H_{y\text{eff}}{}^{\frac{1}{2}}. \quad (6.16)$$

The maximum value is reached for a long needle ($w_x=w_y=\frac{1}{2}$, $w_z=0$)

$$H_{\text{eff}}=H_z+\tfrac{1}{2}M_s.$$

Induced Moments on Domain Boundaries

As the Curie point of a ferromagnetic is approached from higher temperature, the elementary spins spontaneously align and a saturation magnetization **M** develops in a preferred crystallographic direction. Several equivalent lattice directions are available and are usually chosen by the magnetic axes to produce a domain pattern approaching a constellation of minimum free energy.

A large "dynamic" effective field may be produced, as Polder and Smit showed,[28] because of the presence of domain boundaries. Figure 6.2 shows a section through a ferromagnetic body containing 180° domains orientated in the $\pm z$ direction with thin walls in the y-z planes. Application of an oscillatory field in the $\pm x$ direction, that is, perpendicular to the wall, produces a Larmor precession around the H_z field with the sense of rotation opposite in the up-and-down domains. In consequence, the poles in the x direction at the walls appear in phase and, adding to dipole chains, produce the normal shape magnetization factor w_x of the overall body. The poles created in the y direction are in opposition and their demagnetizing effect cancels ($w_{x,\text{eff}}=w_x$, $w_{y,\text{eff}}=0$, $w_z=0$).

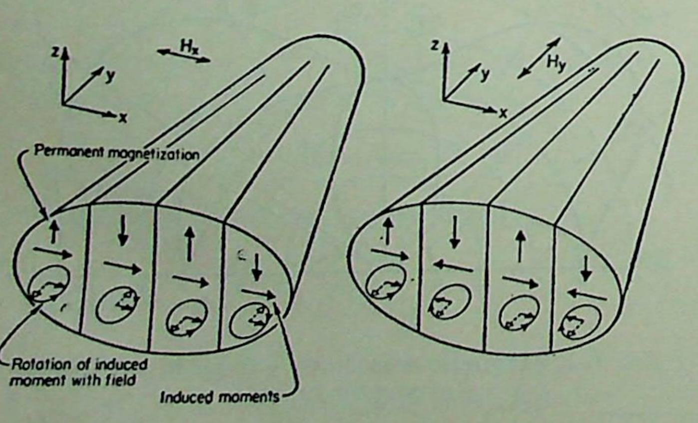


FIG. 6.2. Gyromagnetic rotation creating free poles on domain walls.

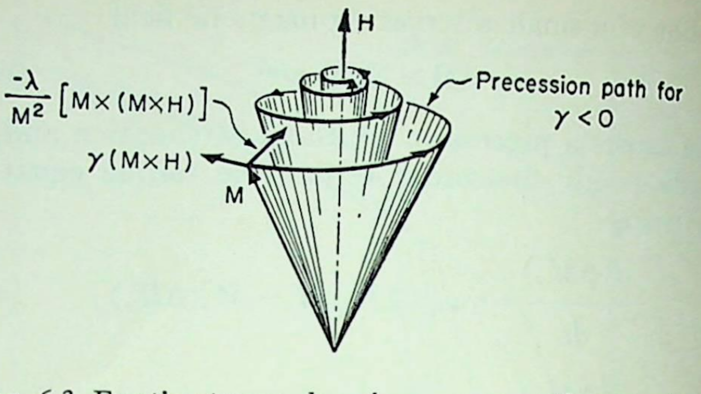


FIG. 6.3. Erecting torque damping gyromagnetic precession.

On the other hand, if the oscillatory field stands in the $\pm y$ direction, the free poles induced on the domain walls pile up in alternating sheets and lead to very strong closing fields in the individual domains ($w_{x\text{eff}}=1$; $w_{y\text{eff}}=w_y$; $w_z=0$). Since the shape factor in the x or y direction can vary between 0 and 1, the effective field for the first mode of excitation can vary between H_z and $H_z{}^{\frac{1}{2}}(H_z+M_z)^{\frac{1}{2}}$, and for the second mode, between $H_z{}^{\frac{1}{2}}(H_z+M_z)^{\frac{1}{2}}$ and (H_z+M_z).

The field action of induced moments invoked by Polder and Smit can be generalized; a system of n types of interacting domains will have n-resonant frequencies as do n-coupled oscillators. A complex domain arrangement pointing at various angles to the applied oscillatory field should therefore produce a broad distribution of Larmor frequencies whose effective fields lie for the most part in the range between the anisotropy field H_a and (H_a+M_z).

In addition to these resonances where the individual moments within each domain precess as a single spin system, the existence of magnetic sublattices in ferrites permits further modes of precession wherein the spin system of one sublattice changes its orientation with respect to the others.[29] The frequency of such modes should lie in the infrared region in most cases.[30] They have never been observed in simple ferrites. However, a resonance of this kind has been induced at microwave frequencies,[31] under the special condition that the sublattice moments were very nearly balanced.

Damping Effects

In real matter, the rotational energy of the Larmor precession will be dissipated by such effects as magnetic dipole radiation, interaction with nearby magnetic moments, with conduction electrons and, through magneto-elastic coupling with lattice vibrations. In consequence, the magnetization vector will spiral back into the z direction. This tendency can be expressed formally in the equations of motion by adding to the

[28] D. Polder and J. Smit, Revs. Modern Phys. 25, 89 (1953).

[29] R. K. Wangsness, Phys. Rev. 91, 1085 (1953); 93, 68 (1954).

[30] J. Kaplan and C. Kittel, J. Chem. Phys. 21, 760 (1953).

[31] R. K. Wangsness, Phys. Rev. 97, 831 (1955); T. R. McGuire, *ibid.* 97, 831 (1955).

rotational torque an erecting torque (Fig. 6.3)

$$\frac{d\mathbf{M}}{dt}=\gamma[\mathbf{M}\times\mu_0\mathbf{H}]-\frac{\lambda}{|\mathbf{M}|^2}[\mathbf{M}\times(\mathbf{M}\times\mathbf{H})]$$
$$\equiv\gamma[\mathbf{M}\times\mu_0\mathbf{H}]+\lambda\left[\mathbf{H}-\frac{(\mathbf{M}\cdot\mathbf{H})\mathbf{M}}{|\mathbf{M}|^2}\right], \quad (6.17)$$

as first introduced by Landau and Lifshitz.[26] Up to the present, no theory has adequately explained why the attenuation is so strong in ferrites; the relaxation times are of the order of 10^{-9} sec. In considering this problem it is convenient to regard the precessional motion of the magnetization vector as a periodic disturbance of the system of spin moments, which leaves all these moments mutually parallel; that is, the wave motion in the spin system has an infinite wavelength or zero wave number **k**. Disturbances of shorter wavelengths may also exist and would have the effect of reducing the over-all magnetic moment of the spin system. Only the wave with $\mathbf{k}=0$ can be excited by a uniform field. Since there is little interaction of this wave by magnetic-elastic coupling with the thermal vibrations of the lattice[32] (the bulk of which have wave numbers of the order of 10^9 m^{-1}), it has been difficult to explain the attenuation as a direct energy transfer process. It has been suggested[32] that energy is first transferred from the $\mathbf{k}=0$ wave to waves of higher **k** value and thence to the lattice, but this process itself has remained unexplained, since in a system of infinite extent much more energy is required to excite a wave of high **k**. This is because, when neighboring spins are moved out of

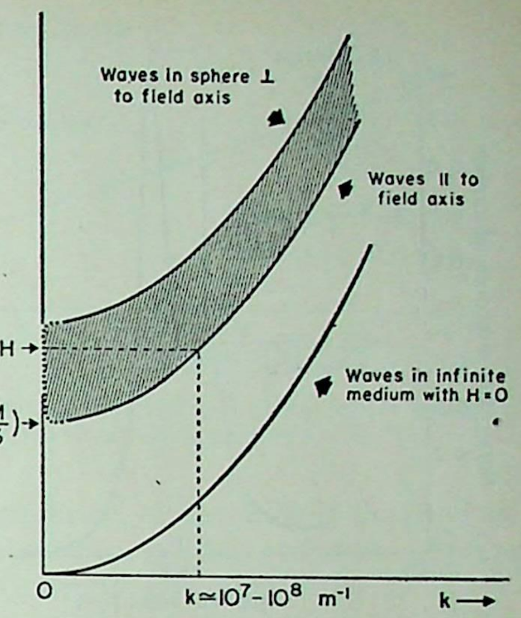


Fig. 6.5. Energy of spin waves propagating in sphere.

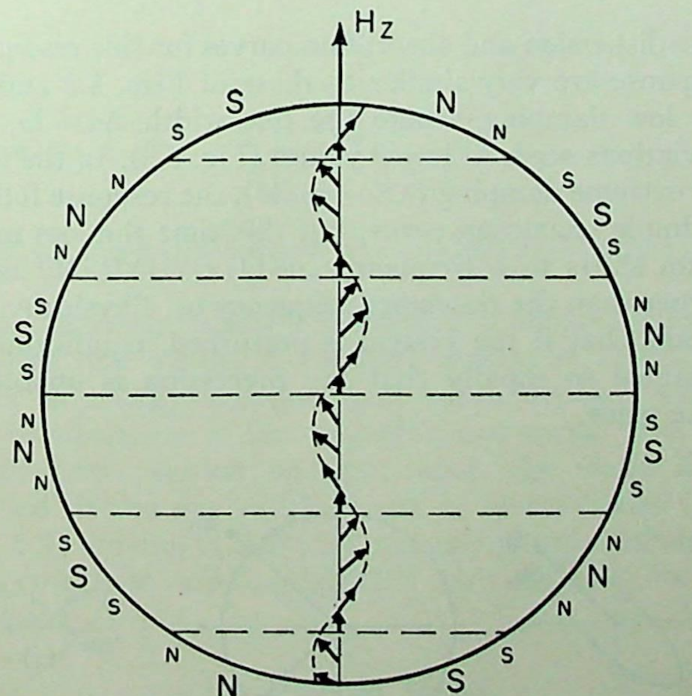


FIG. 6.4. Dynamic pole distribution on sphere caused by high-order spin wave.

alignment, exchange energies enter of the form

$$\mathcal{E}_k=\gamma\mu_0 H\hbar+Xk^2, \quad (6.18)$$

where X is a constant of the order of 10^{-40} joule m^2.

Lately, it has been realized[33] that for a magnetic system of finite extent this energy expression has to be modified considerably. For example, a disturbance of large wave number in a sphere, oriented along the external field direction, induces alternate N and S poles on the surface of the sphere (Fig. 6.4), which, in contrast to the case of $\mathbf{k}=0$, effectively cancel each other's influence at a distance large to their separation. Therefore the precessing dipoles of this wave mode contribute magnetization components only in the z direction, hence the demagnetizing factors are $w_z=\frac{1}{3}$, $w_x=w_y=0$, and the energy of the wave is (Fig. 6.5)

$$\mathcal{E}_k=\hbar\gamma\mu_0\left(H_z-\frac{M}{3}\right)+Xk^2. \quad (6.19)$$

For wave propagation at an angle θ to the field axis in a spheroid

$$\mathcal{E}_k=[\hbar\gamma\mu_0(H_z-N_zM)+Xk^2]^{\frac{1}{2}}$$
$$\times[\hbar\gamma\mu_0(H_z-N_zM+M\sin^2\theta)+Xk^2]^{\frac{1}{2}}. \quad (6.20)$$

For very low **k** values, the situation is more complicated and has been treated theoretically by Walker.[34] Such modes have been excited by nonuniform fields[35] (Fig. 6.6). It is clear from (6.20) and Fig. 6.5 that many modes with comparatively large $\mathbf{k}(\simeq10^{7\to8}[m^{-1}])$ may have approximately the same energy as the $\mathbf{k}=0$ wave.

[32] C. Kittel and E. Abrahams, Revs. Modern Phys. **25**, 233 (1953).

[33] Clogston, Suhl, Walker, and Anderson, Phys. Rev. **101**, 903 (1956).

[34] L. R. Walker, Phys. Rev. **105**, 390 (1957).

[35] P. A. Miles, thesis, London, 1954 (unpublished); White, Solt, and Mercereau, Bull. Am. Phys. Soc. Ser. II, **1**, 12 (1956); J. F. Dillon, Jr., *ibid.* Ser. II, **1**, 125 (1956); R. L. White and I. H. Solt, Jr., Phys. Rev. **104**, 56 (1956).

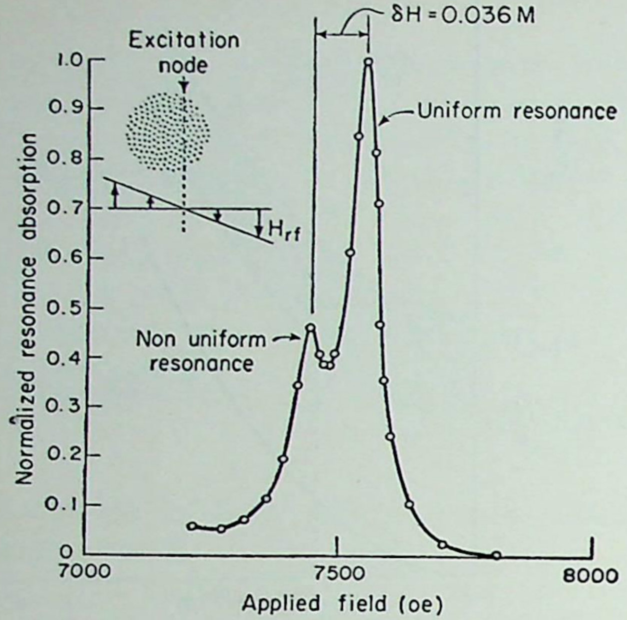


FIG. 6.6. Magnetic resonance doubling at 24 kmc in a nickel ferrite crystal sphere produced by nonuniform excitation (Miles[35]).

The fact that structural inhomogeneities in the lattice may provide coupling between these waves offers some hope for the explanation of resonance damping in ferrites.

The resonance frequency, hence energy, of the uniform mode depends on the shape of the specimen. It follows that the degree of degeneracy of the spin-wave system, and with it the resonance damping, may also be shape-dependent. Thus the damping for a rod magnetized along its axis should be more than for a sphere, because more spin-wave modes of approximately equal energy can be excited in the rod. Such effects have not yet been reported. The spin-wave theory predicts[36] in addition that the uniform precession mode, directly excited, may become unstable at large amplitudes of oscillation. The reduction of resonance absorption in metals and ferrites at high levels of rf excitation observed by Bloembergen and co-workers[37] is believed to be due to this instability.

We will use the Landau-Lifshitz formalism for the future discussion, despite the fact that the mechanism outlined above would not conserve the length of the vector **M**. The damping factor will therefore be considered as a parameter which may depend on both **H** and **M**.

The solution of (6.17) is

$$M_x = M_0 e^{(j\omega_0-\alpha)t}$$
$$M_y = M_0 e^{(j\omega_0-\alpha)t+\pi/2} \tag{6.21}$$
$$M_z = M\left\{1-\left(\frac{M_0}{M}\right)^2 e^{-2\alpha t}\right\},$$

with $\alpha=(\lambda H/M)$. By introducing the erecting torque, the oscillators in the x-y plane remain out of phase by 90° in space and time but are attenuated. The resonance frequency ω_0 is still identical to the Larmor frequency of (6.7).

Corresponding to (6.10), which relate to the application of a small oscillatory field, we now have

$$\Delta M_x = a'(\Delta H_x) - jb'(\Delta H_y)$$
$$\Delta M_y = a'(\Delta H_y) + jb'(\Delta H_x), \tag{6.22}$$
$$\Delta M_z = 0$$

where

$$a' \equiv \frac{\omega_0\omega' + \lambda\alpha + j\omega\lambda}{\omega_0^2 + (j\omega+\alpha)^2}, \quad b' \equiv \frac{\omega\omega'}{\omega_0^2 + (j\omega+\alpha)^2}$$

and $\omega' = -\gamma\mu_0 M_z$.

For very low frequencies in relation to ω_0, the Δ**M** vector in the plane oscillates practically in the direction of and in phase with the driving field; for very high frequencies ($\omega \gg \omega_0$), the magnetization vector is essentially out of phase by 90° in space and 180° in time; at the Larmor frequency itself, the Δ**M** vector describes a circular motion in the x-y plane (Fig. 6.7) with a temporal phase shift of 90° with respect to the driving field ΔH. Energy is continuously absorbed from the exciting field and transferred to the surrounding medium via the damping process.

In comparison with the case of undamped motion (6.12), the diagonal susceptibility terms now have the form

$$\frac{\chi_m^*}{\chi_{ms}'} = \frac{\omega_0^2+\alpha^2+j\omega\alpha}{\omega_0^2+\alpha^2-\omega^2+j\omega 2\alpha}. \tag{6.23}$$

The dispersion and absorption curves for this resonance response are very similar to those of Figs. 4.2 and 4.4 for low damping, where the line width $\Delta\omega = 2\alpha$, but deviations occur at low Q values (Fig. 6.8). In the limit for extreme damping ($\lambda \gg -\gamma\mu_0 M$), the response follows a simple relaxation curve, but this time the loss maximum shifts to a frequency $\omega = (1/\tau) = (\lambda H/M)$ much higher than the resonance frequency ω_0. Physically this means that if the system is perturbed, equilibrium is regained so rapidly that the precession is unable to take place.

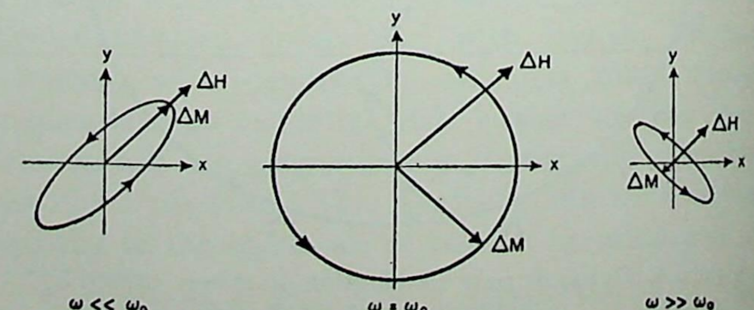


FIG. 6.7. Equilibrium precession of magnetization vector in small oscillatory field.

[36] P. W. Anderson and H. Suhl, Phys. Rev. **100**, 1788 (1955).
[37] N. Bloembergen and S. Wang, Phys. Rev. **93**, 72 (1954); R. W. Damon, Revs. Modern Phys. **25**, 239 (1953).

If the factor λ is a physical constant of the material, the Q value of the resonance is independent of the frequency, since

$$Q=\frac{\omega_0}{2\alpha}=-\frac{\gamma\mu_0 H_{res}}{2\lambda H_{res}}\cdot M=-\frac{\gamma\mu_0 M}{2\lambda}. \qquad (6.24)$$

Thus if the resonance is traced by varying the applied field $\mathbf{H}$ at a fixed frequency ω, the resonance line width ΔH should be proportional to ω. In ferrite single crystals, line widths measured in the microwave region are found to vary linearly with ω, but not proportional to ω (Fig. 6.9). Thus, no unique frequency-independent constant λ can be used to describe such resonances in ferrites. In the frequency region 10^{10} to 4×10^{10} cps, at least, we can take care of this fact in a formal way by introducing a further damping factor δ in the form

$$\partial\mathbf{M}/\partial t=\gamma\mu_0(\mathbf{M}\times\mathbf{H}) -\left(\frac{\lambda}{M^2}+\frac{\delta}{MH}\right)(\mathbf{M}\times\mathbf{M}\times\mathbf{H}). \qquad (6.25)$$

The attenuation factor α of (6.21) now becomes

$$\alpha=\frac{\lambda H}{M}+\delta \qquad (6.26)$$

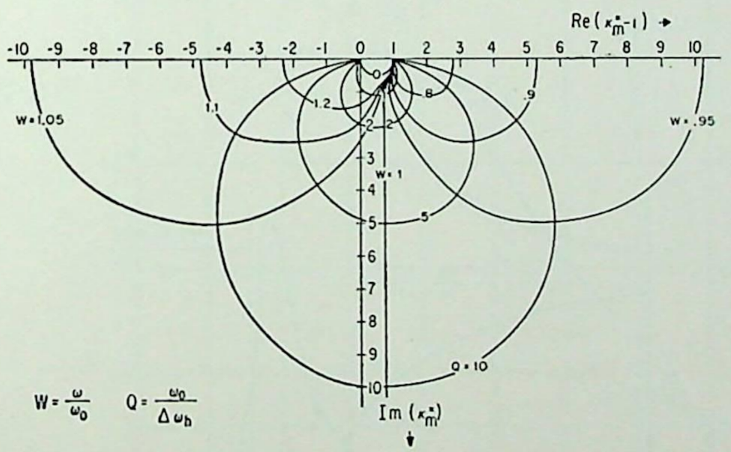


FIG. 6.8. Resonance response with Landau-Lifshitz damping.

with the contribution δ describing a frequency-independent line width. A damping process of this form gives an energy loss per cycle which depends solely on the deviation of the magnetization vector from its equilibrium position and not upon the static field applied. If the line widths found in nickel ferrite (see Fig. 6.9) are due to damping, the appropriate values of the constants are $\delta=3.2\times10^8$ rad/sec $\lambda=1.3\times10^8$ rad/sec.

(c) Domain Walls

Adjacent areas in a crystal, where the spin arrays are aligned along different axes of easy magnetization, are separated by transition regions (domain walls) in which the spins change direction gradually (Fig. 6.10).

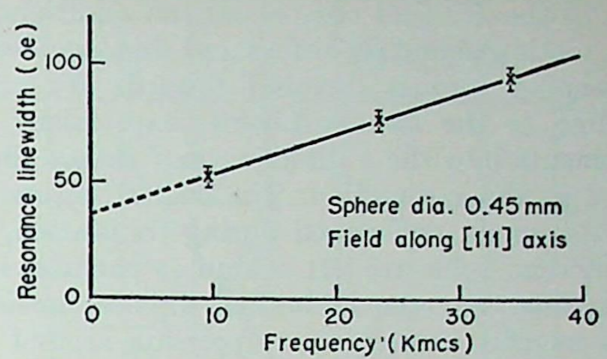


FIG. 6.9. Resonance line width of nickel ferrite single crystal at microwave frequencies (Miles[35]).

The precise structure of the wall is determined by a compromise between opposing forces of magneto-crystalline anisotropy trying to align all spins along the preferred axes (hence to minimize the thickness of the transition layer) and exchange energy, forcing spins into parallel alignment (hence expanding the wall in order to minimize the misalignment of neighboring spins). The resultant energy stored in a wall may be shown[21,38] to be proportional to the product $(AK_1)^{\frac{1}{2}}$, and its thickness of the order of $(A/K_1)^{\frac{1}{2}}$ where K_1 is the first-order anisotropy constant and A an exchange-energy factor usually of the order of 10^{-10} joule/m. Estimates of wall energies for known ferromagnetics lie in the range 10^{-4} to 10^{-2} joule/m² for walls of 10^2 to 10^4 angstroms thickness.

The motion of domain walls has been treated in the following way[26,39]: the application of an external field H_z produces a torque on the magnetic spins in the wall proportionally to $\sin\theta$. Since the spins behave as gyroscopes, there tends to develop a precession around the z axis carrying the spins out of the y-z planes. Because the spins are coupled and staggered, an actual precession cannot take place but a wall deformation results, producing a magnetization in the x direction. This component M_x creates a closing field

$$H_x=-M_x \qquad (6.27)$$

because div $B_x=0$ (div B_y and div B_z are, *a priori*, zero for a wall of infinite sidewise extension). The torque exerted by H_z, and hence the wall distortion and the M_x components, are a maximum in the middle of the wall, where the closing field H_x may be appreciably larger than the applied field H_z.

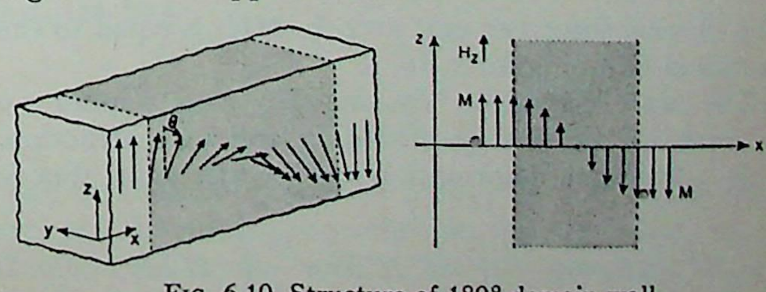


FIG. 6.10. Structure of 180° domain wall.

[38] C. Kittel, Revs. Modern Phys. 21, 541 (1949).
[39] W. Döring, Z. Naturforsch. 3A, 373 (1948); R. Becker, J. phys. radium 12, 332 (1951); Rado, Wright, and Emerson, Phys. Rev. 80, 273 (1950).

Due to the H_x field component, an additional precession motion about H_x ensues and this, coupled with the damping torques directed towards H_x and H_z (according to the Landau-Lifshitz expression) erects the moments into the z direction and causes the wall to move in the x direction. The wall structure is assumed to remain unchanged during translation, while the individual spins are left behind as particles in any wave motion. For the individual dipoles the instantaneous angular velocity for precession around the x axis is

$$\omega_0 \equiv \frac{d\theta}{dt} = -\gamma\mu_0 H_x, \tag{6.28}$$

and determines the translation velocity $\boldsymbol{v}$ (Fig. 6.10) as

$$-\frac{d\theta}{dt} = \frac{d\theta}{dx} \cdot \boldsymbol{v}; \tag{6.29}$$

hence

$$H_x = \frac{\boldsymbol{v}}{\mu_0\gamma} \frac{d\theta}{dx}. \tag{6.30}$$

In moving the wall a unit distance, the potential energy of the system is reduced by the reversal from antiparallel to parallel of moment $2M$ per unit area of wall. The associated energy $2\mu_0 M H_z$ must be transferred to the lattice by the damping mechanism operating on the re-orientating dipole moments in the wall. According to (6.17), this mechanism gives a dissipation rate $\mu_0 H \cdot (dM/dt) \simeq \mu_0 \lambda H_x^2$, since for low damping $H_x \gg H_z$ in the center of the wall. The consequence of the introduction of the Landau-Lifshitz term is to give a rate of damping which increases with wall distortion, and hence with wall velocity.

The equilibrium velocity $\boldsymbol{v}$ is given by the balance of the rates of energy creation and dissipation

$$2MH_z\mu_0\boldsymbol{v} = \int_{-\infty}^{\infty} \mu_0\lambda H_x^2 dx = \frac{\boldsymbol{v}^2\lambda}{\gamma^2\mu_0} \int_{-\infty}^{\infty} \left(\frac{d\theta}{dx}\right)^2 dx,$$

whence (6.31)

$$\boldsymbol{v} = \frac{2\gamma^2\mu_0^2 M}{\lambda \int_{-\infty} \left(\frac{d\theta}{dx}\right)^2 dx} \cdot H_z = \frac{2\mu_0 M H_z}{\beta}.$$

The driving force per unit area $2\mu_0 M H_z$ is equal to the frictional damping force $\beta\boldsymbol{v}$.

The wall distortion causing H_x also implies an increase in the energy stored in the wall structure. This additional deformation energy per unit area is

$$\Delta\mathcal{E}_w = -\frac{\mu_0}{2}\int_{-\infty}^{\infty} M_x H_x dx = \frac{\mu_0}{2}\int_{-\infty}^{\infty} H_x^2 dx$$
$$= \frac{\boldsymbol{v}^2}{2\gamma^2\mu_0}\int_{-\infty}^{\infty}\left(\frac{d\theta}{dx}\right)^2 dx. \tag{6.32}$$

The coefficient of $\boldsymbol{v}^2/2$ may be interpreted as an effective wall mass

$$m = \frac{1}{\gamma^2\mu_0}\int_{-\infty}^{\infty}\left(\frac{d\theta}{dx}\right)^2 dx = \frac{\beta}{\lambda}. \tag{6.33}$$

This description of wall motion tacitly assumes that the equation of motion for individual spins can be handled as if all the magnetization vectors would execute their precession with equal amplitude and phase. Actually, the introduction of a wall distortion that moves past lattice imperfections suggests that exchange coupling effects may become of importance.[40] Furthermore, a wall region may differ in its properties from the bulk material (Sec. 7); hence damping processes effective in domain precession and domain-wall motion may not be identical.

When a wall is stabilized in position by the presence of impurities, cavities, stress patterns, etc., an elastic restoring force of the type $F_x = -kx$ exists for small displacements. Consequently, the equation of motion of the trapped wall may be written as that of an harmonic oscillator of the mass m subjected to frictional damping, i.e.,

$$m\ddot{x} + \beta\dot{x} + kx = 2\mu_0 M H_z. \tag{6.34}$$

If a periodic field $H_z = H_z e^{j\omega t}$ is applied, a susceptibility contribution stemming from the oscillating walls is observed,

$$\chi_{mz}^* \equiv \frac{\Delta M_z}{H_z} = \frac{2Mx}{H_z} n. \tag{6.35}$$

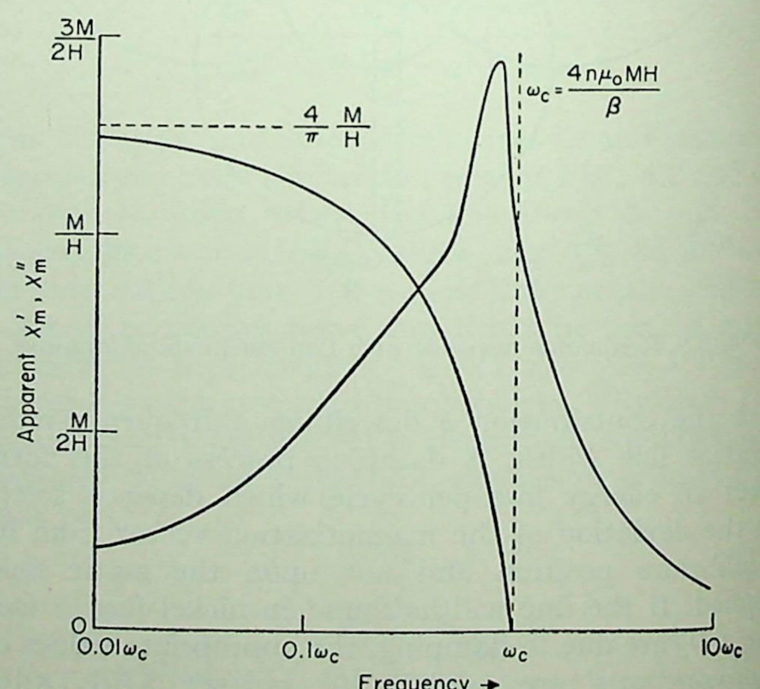


FIG. 6.11. Contribution to magnetic dispersion and loss by migrating domain walls (fundamental component derived from harmonic analysis).

[40] Exchange effects have been observed in a ferromagnetic resonance experiment in a metal, where a pronounced skin effect caused the exciting field, and with it the precession motion, to be rapidly attenuated with depth (W. S. Ament and G. T. Rado, Phys. Rev. **94**, 1411 (1954); **97**, 1558 (1954).

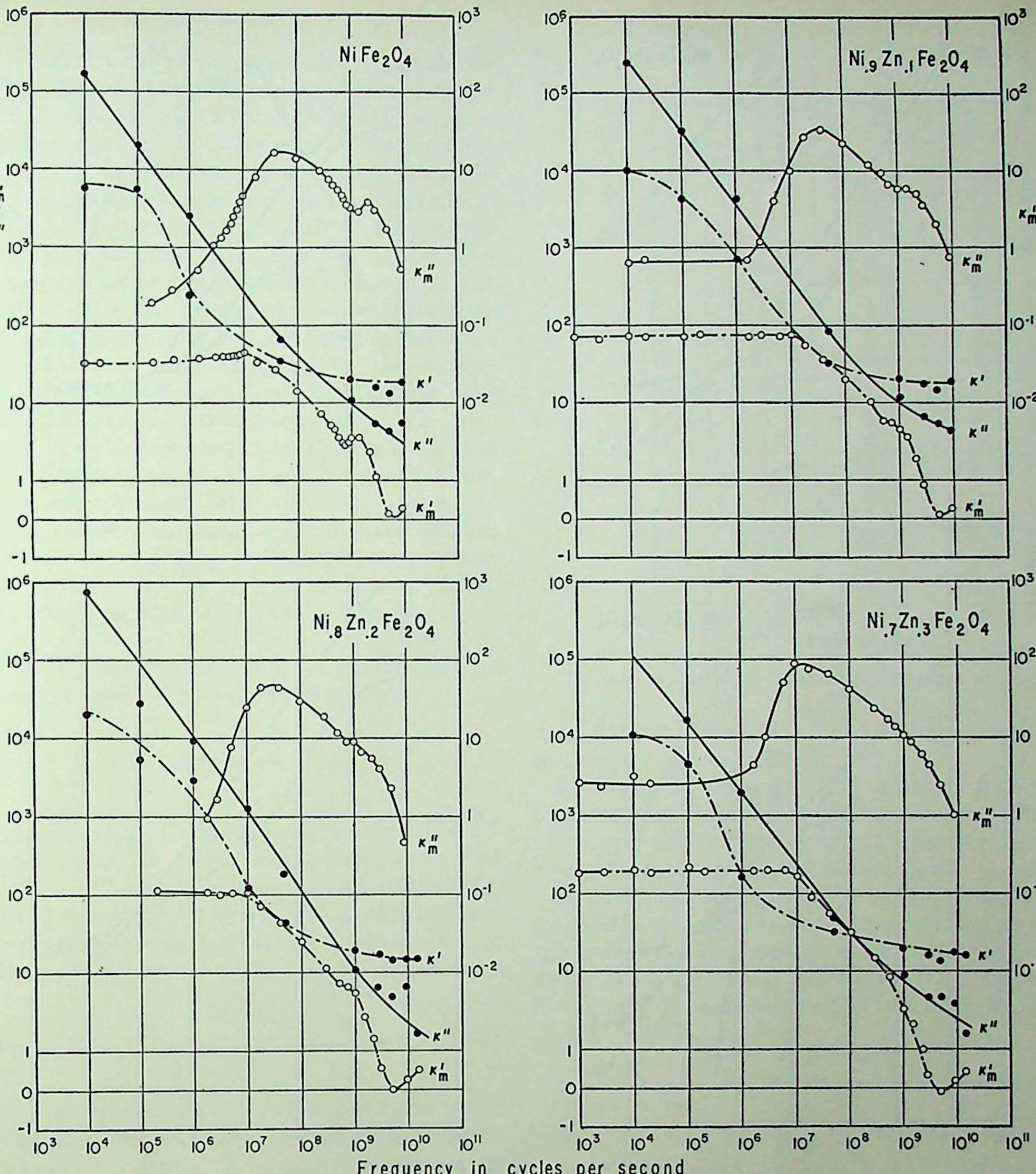


FIG. 7.1. Dielectric spectra of nickel-zinc ferrites from 10^3 to 10^{11} cps.

Here $2Mx$ represents the moment reversed for one wall displaced by x, and n the number of walls per unit length, in the x direction. Hence

$$\chi_{mz}^* = \frac{n4\mu_0 M^2/m}{\omega_0^2 - \omega^2 + j\omega\lambda}, \tag{6.36}$$

with $\omega_0^2 = k/m$. The ratio of dynamic to static susceptibility is therefore

$$\frac{\chi_{mz}^*}{\chi_{ms}} = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + j\omega\lambda} = \frac{k}{k - m\omega^2 + j\omega\beta} \tag{6.37}$$

where $\chi_{ms} = 4n\mu_0 M^2/k$. In the case of high damping ($\beta \gg m\omega$), this resonance response approaches a relaxation form

$$\frac{\chi_{mz}^*}{\chi_{ms}} \simeq \frac{k}{k + j\omega\beta} = \frac{1}{1 + j\omega\tau}, \tag{6.38}$$

with the relaxation time $\tau = (\beta/k) = (\lambda/\omega_0^2)$. If the walls are not bound to equilibrium positions but move through the crystal with a velocity given by (6.31), the susceptibility becomes

$$\chi_{ms}^* = -j\left(\frac{4n\mu_0 M^2}{\beta\omega}\right), \tag{6.39}$$

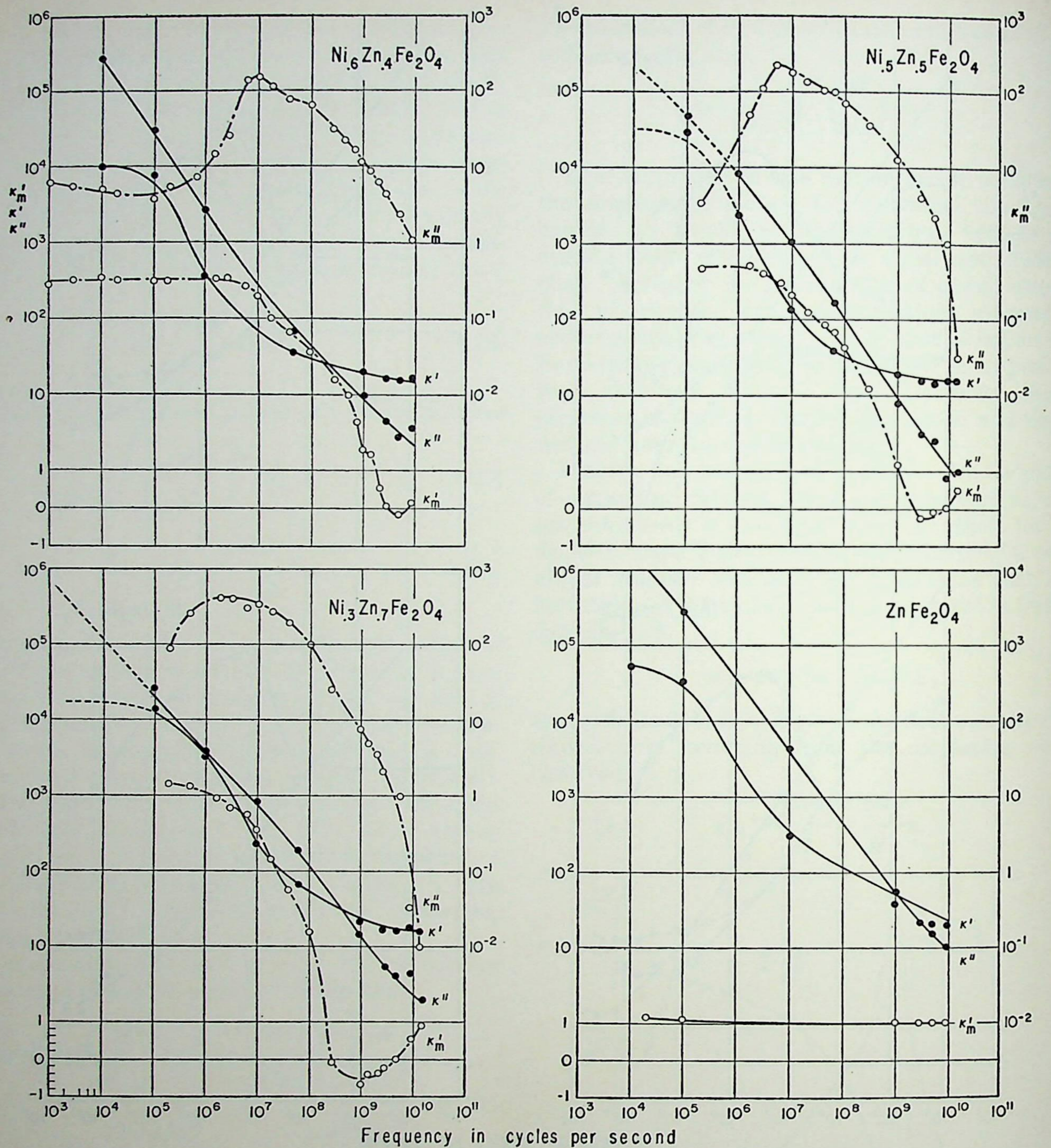


FIG. 7.2. Dielectric spectra of nickel-zinc ferrites from 10^3 to 10^{11} cps.

that is, a pure magnetic loss which varies with frequency in the same way as an electric loss due to charge migration. If this wall motion is interrupted at the critical frequency $\omega_c = 4n\mu_0 MH_z/\beta$ by the walls being swept out of the crystal, the magnetic response is no longer linear. The harmonic analysis of this new response at frequencies below ω_c has fundamental components both in and out of phase with the exciting field (*Fig. 6.11*). It can be seen that a distribution of critical frequencies can produce a dispersion spectrum similar to a relaxation, but this will be accompanied by an abnormally high and steep loss characteristic. The introduction of a finite coercive field H_c produces a nonlinear response at all frequencies and an additional frequency-dependent loss component below the appropriate critical frequency.

7. MAGNETIC SPECTRA OF FERRITES

The preceding discussion foresees three dispersion mechanisms: migration of domain walls, oscillation of domain walls around equilibrium positions, and gyromagnetic rotation of the magnetization vectors within domains. The migration of walls causes a loss similar to conduction phenomena, and possibly a relaxation-like spectrum; wall oscillation and gyromagnetic rotation, if not overdamped, produce resonance spectra.

Wall or Spin Resonance?

The frequency-response characteristics of the nickel-zinc ferrite series at room temperature (Figs. 7.1 and 7.2), as well as those of other ferrites, give the impression that there are two main dispersion regions showing some resonance character. The first is marked by a rapid increase in magnetic loss combined with a slight rise in the κ_m' characteristic in the range of 10^6 to 10^7 cps; the second by a loss peak and a minimum of κ_m' between 10^9 and 10^{10} cps. Rado and co-workers[41] advanced the interpretation that the upper and lower resonance regions are caused by spin and by domain-wall resonance, respectively, while scientists at the Philips Research Laboratories[42] are inclined to attribute both regions in cubic ferrites to spin resonance.

It cannot be postulated that the frequency location of the resonance region gives an *a priori* decision about the mechanism involved. The two main pieces of experimental evidence supplied by these characteristics are the static susceptibility χ_{ms} and the resonance frequency as identified by a loss maximum. These quantities are theoretically interrelated (Sec. 6).

For *wall resonance*, where $\omega_0=(k/m)^{\frac{1}{2}}$ and $\chi_{ms}=4C\mu_0M^2/k$ (6.36) with the factor C given by the number of walls and their relative orientation,

$$\omega_0^2\chi_{ms}=\frac{4C\mu_0M^2}{m}. \tag{7.1}$$

By introducing typical experimental values: $M=2.5\times10^5$ amp/m, $\chi_{ms}=30$, 10^4 walls per meter, and a wall mass as calculated for nickel ferrite (9×10^{-10} kg/m²), we obtain a resonance frequency $\omega_0/2\pi=55$ Mc/sec. Over-damping of the wall leads to a loss-maximum frequency $\omega_d=(k/\lambda m)$ and

$$\omega_d\chi_{ms}=\frac{4C\mu_0M^2}{\lambda m}. \tag{7.2}$$

With a typical damping factor, taken (with the reservations discussed in Sec. 6b) from magnetic resonance

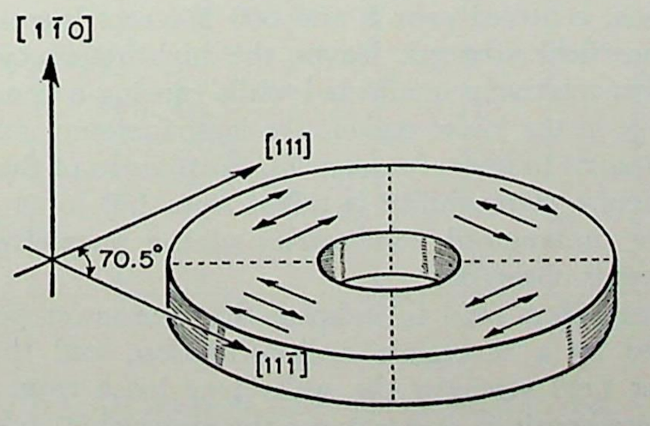


FIG. 7.3. Toroid of nickel ferrite single crystal cut in (110) plane.

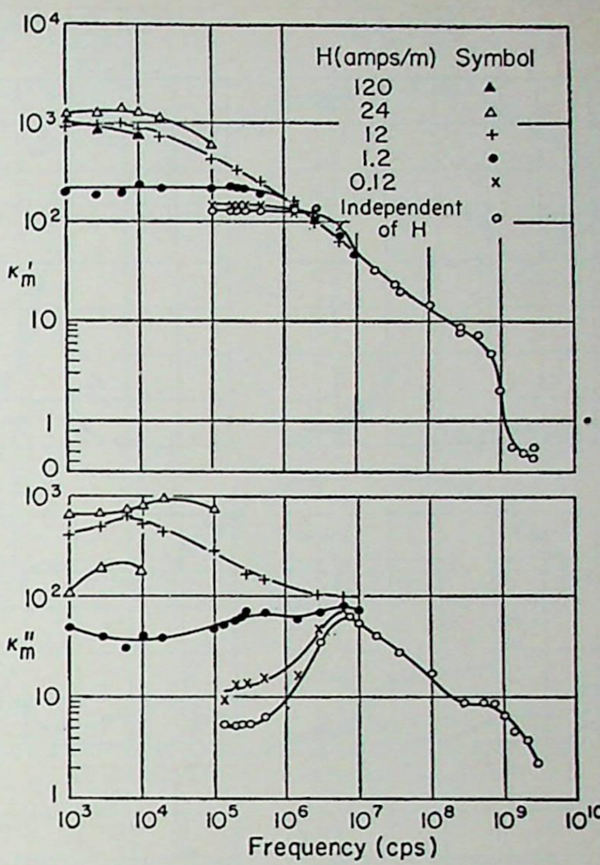


FIG. 7.4. Magnetic spectrum of nickel ferrite single crystal.

data ($\lambda\simeq1\times10^9$ rad/sec), the loss-maximum frequency is lowered to $\omega_d/2\pi\simeq20$ Mc/sec.

Deriving the corresponding relations for *spin resonance* requires knowledge of the effective (inner) field which acts on the spin system; it depends on anisotropy energy and on shape and domain-walls factor. If the inner field stands in the z direction ($H_{z\mathrm{eff}}$), a small magnetic applied in the x or y direction may see different susceptibilities because the inner field does not necessarily have rotational symmetry (i.e., the restoring forces may be different in the two directions). Hence (6.16)

$$\chi_{(ms)x}=\frac{C'M}{H_{x\mathrm{eff}}}$$

$$\chi_{(ms)y}=\frac{C'M}{H_{y\mathrm{eff}}} \tag{7.3}$$

$$\omega_0=-\gamma\mu_0(H_{x\mathrm{eff}}H_{y\mathrm{eff}})^{\frac{1}{2}}.$$

Thus

$$\omega_0\chi_{(ms)x}=-\gamma\mu_0C'M\left(\frac{H_{y\mathrm{eff}}}{H_{x\mathrm{eff}}}\right)^{\frac{1}{2}}; \tag{7.4}$$

in the case of axial symmetry ($H_{x\mathrm{eff}}=H_{y\mathrm{eff}}$) the product is simplified to

$$\omega_0\chi_{ms}=-\gamma\mu_0C'M. \tag{7.5}$$

The factor C' takes account of the spatial distribution of the spin axes of the various domains. The experimental values given above, plus a random distribution for the various spin-axis directions ($C'=\frac{2}{3}$), lead to a Larmor frequency $\omega_0/2\pi=180$ Mc/sec. If the spin

[41] G. T. Rado, Revs. Modern Phys. 25, 81 (1953).
[42] J. L. Snoek, Physica 14, 207 (1948); D. Polder and J. Smit, Revs. Modern Phys. 25, 89 (1953).

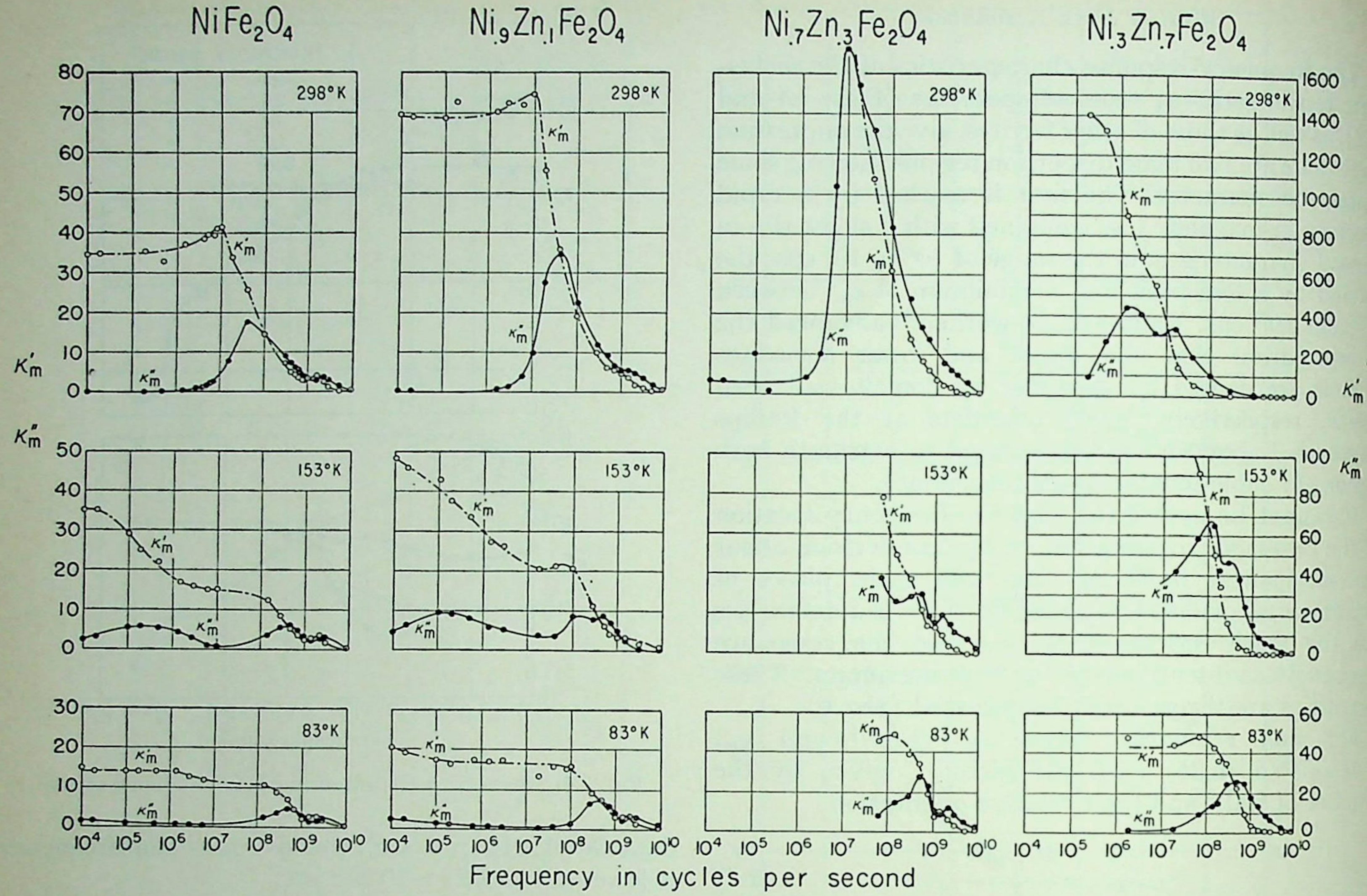


FIG. 7.5. Temperature dependence of magnetic spectra of nickel-zinc ferrite ceramics.

resonance becomes overdamped ($\lambda > |\gamma\mu_0 M|$), the loss maximum occurs at a frequency

$$\omega_d = \alpha = \frac{\lambda H_{\text{eff}}}{M}, \tag{7.6}$$

hence

$$\omega_d \chi_{ms} = C'\lambda. \tag{7.7}$$

However, with $\lambda = 1 \times 10^9$ rad/sec, as above, the dispersion would retain its resonance character and the loss-peak frequency would remain practically unchanged.

The domain-wall resonance in our example lies at a frequency somewhat lower than that of the spin resonance, but due to the arbitrary choice of the number of walls present, there is no clear order-of-magnitude distinction. When dealing with unknown domain arrays and multicrystalline structure of ceramics, the whole calculation rests on shaky grounds. Additional information can be provided by measurements on single crystals, by temperature- and field-strength effects and by a careful analysis of the influence of cation substitution.

Single-Crystal Measurements

The only single crystal available in the nickel-zinc ferrite series was a nickel-ferrite crystal grown by flame fusion.[43] It was cut in the shape of a toroid, oriented with a (110) plane parallel to the toroid plane (Fig. 7.3); the two [111] directions in this plane are directions of easy magnetization at room temperature. The cylindrical shape was a compromise between the ideal diamond shape with a simple domain pattern for quasi-static studies,[44] and the need to take measurements over a wide frequency range.

At low field strengths the frequency-response curve (Fig. 7.4) shows a general similarity to that of the ceramic, but the loss peaks are higher and more widely separated. It appears that there are two dispersion regions, centered near 8 and 600 Mc/sec. Increase of the ac field strength leaves the high-frequency dispersion relatively unaffected while causing a complete change in the lower region: the loss maximum flattens and moves to lower frequencies, the plateau of the low-frequency permeability is raised from 130 to *ca* 1300, and a pronounced nonlinearity of the magnetization process is observed.

These facts tend to indicate that the lower peak is caused by a reversible wall oscillation, and that at higher field strength the walls tear loose from their anchored position and traverse the material in widening swings.

[43] Supplied through courtesy of Linde Air Products Company.
[44] Galt, Andrus, and Hopper, Revs. Modern Phys. 25, 93 (1953).

Temperature Effects

Lowering the temperature of the ceramics produces a very striking effect (Figs. 7.5 and 7.6): a broad relaxation spectrum splits off from the main dispersion region and moves rapidly to lower frequencies; simultaneously, two loss peaks become clearly discerned in the range from 10^8 to 10^{10} cps. A corresponding change occurs in the magnetic spectrum of the single crystal (Fig. 7.7).

A detailed study of the relaxation spectrum[45] shows that it stems from some process requiring an activation energy U. The frequency ω_m at which the maximum of the loss tangent is traversed, varies with temperature as

$$\omega_m = \omega_0 e^{-U/kT}, \tag{7.8}$$

and a Cole-Cole plot places all measurements on circular arcs with centers below the real axis (Fig. 7.8).‡

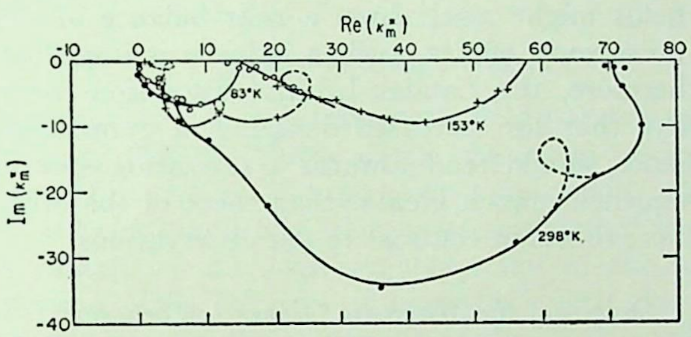


FIG. 7.6. Cole-Cole plot for data of Fig. 7.5 ($Ni_{0.9}Zn_{0.1}Fe_2O_4$).

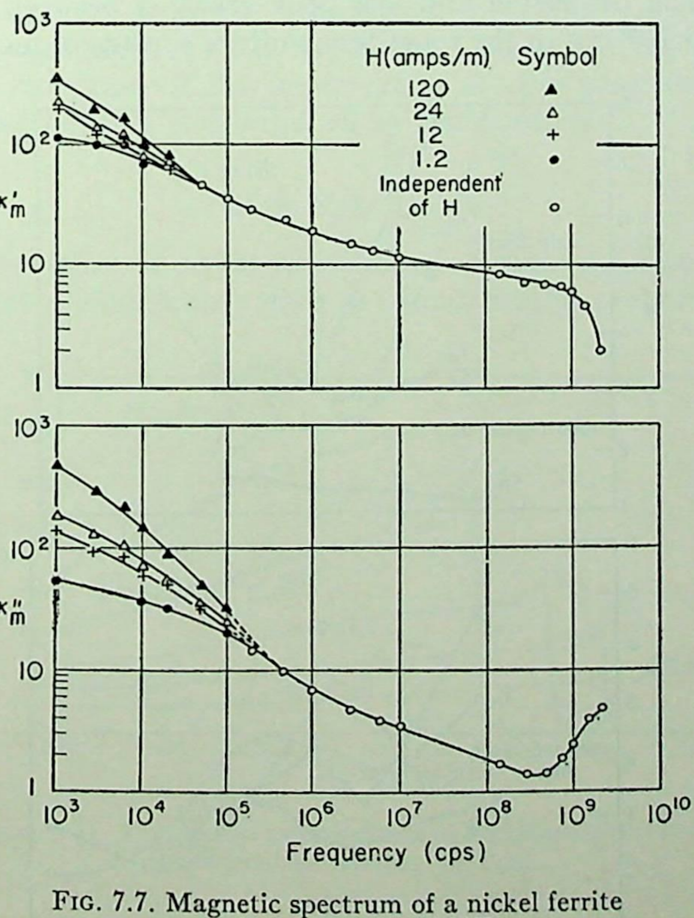


FIG. 7.7. Magnetic spectrum of a nickel ferrite single crystal at 83°K.

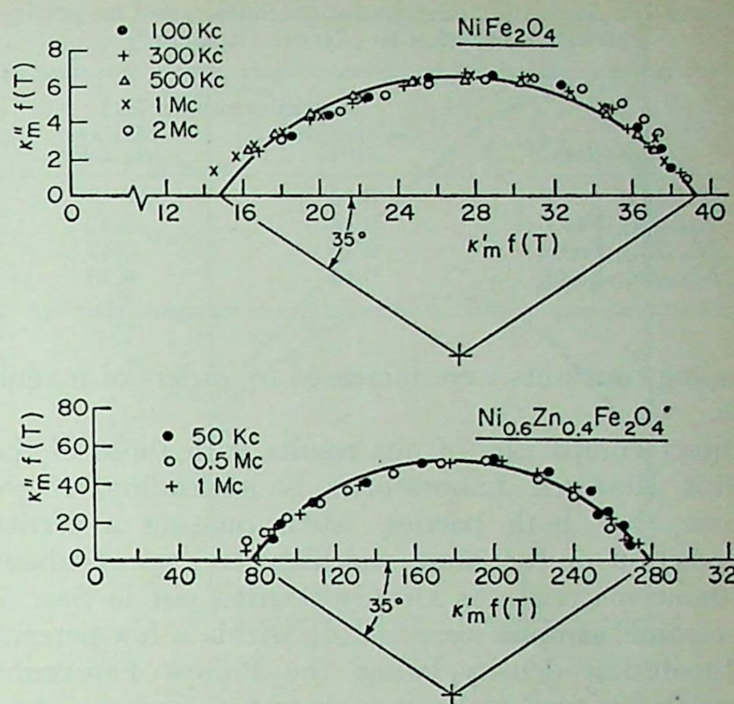


FIG. 7.8. Cole-Cole plot of relaxation spectra for nickel-zinc ferrite ceramics (Epstein[45]).

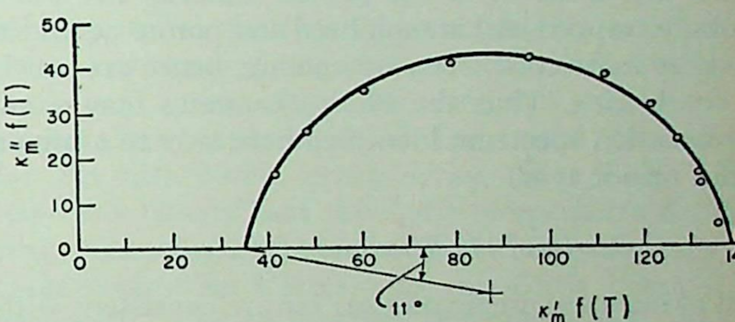


FIG. 7.9. Cole-Cole plot of relaxation spectrum for nickel ferrite crystal (Epstein[45]).

The data for ceramics are fitted by a distribution of time constants extending over two decades. This is not surprising because individual crystallites differ in composition and state of oxidation, in strain and environment. This is confirmed by measurements on the nickel ferrite single crystal which give a distribution of time constants about thirty times narrower than that for the corresponding ceramic (Fig. 7.9).

Activation energies obtained from these data (Table 7.1) agree well with those of conductivity measurements, and with previous findings of Wijn.[23,46] Temperature has a very striking influence on the switching of memory cores made from these materials: while at room temperature the switching takes place in less than 10^{-6} sec, Epstein[45] observed time constants of seconds at 77°K. This corresponds to the damping effects seen by Galt[47] in domain-wall motion studies on single crystals, where at these low temperatures the apparent

---

[45] D. J. Epstein, Sc.D. thesis, Massachusetts Institute of Technology, May, 1956.

‡ The relaxation spectrum can be traversed either by varying the frequency at fixed temperature, or by temperature variation at fixed frequency. In the latter case the range of the relaxation dispersion ($\kappa_{ms}' - \kappa_{m\infty}'$) is found to vary with temperature. The function $f(T)$ appearing in Fig. 7.8 is an empirical factor introduced to compensate for this change.

[46] H. P. J. Wijn and H. van der Heide, Revs. Modern Phys. 25, 98 (1953).

[47] J. K. Galt, Bell System Tech. J. 33, 1023 (1954); *ibid.* 34, 439 (1955).

TABLE 7.1. Activation energies for conduction and magnetic relaxation in nickel-zinc ferrite (Epstein[45]).

Composition	Activation energies (ev) Conductivity 170°K	Magnetic relaxation
$NiFe_2O_4$	0.14	0.17
$Ni_{0.8}Zn_{0.2}Fe_2O_4$	0.14	0.12
$Ni_{0.6}Zn_{0.4}Fe_2O_4$	0.16	0.17
$Ni_{0.3}Zn_{0.7}Fe_2O_4$	0.14	0.13

damping constants were increased by orders of magnitude.

Direct comparison of our results with those of the Philips Research Laboratories is misleading, if we assume that both parties, when quoting a ferrite composition of the same stoichiometry, speak about the same material. As already pointed out in Sec. 5, our ceramic samples were fired to within a few percent of theoretical density, while the Philips Feroxcube materials are fired at lower temperature and maintain a porous structure. The consequences for domain-wall motion are decisive: in the porous material the walls are easily trapped and immobilized and porous ceramics re-oxidize more completely on cooling, hence are much less conducting. Thus the Philips ceramics may show the relaxation spectrum identified here only to a minor degree, or not at all.

Origin of the Relaxation Spectrum

All of our measurements thus far are consistent with the following interpretation of the relaxation spectrum.

In slightly reduced ferrites Fe^{2+} and Fe^{3+} cations are built in at octahedral lattice sites and can easily exchange electrons, as is known from studies on magnetite. Transfer of electrons also represents a transfer of magnetic spins. The domain-wall array in a crystal corresponds to some minimum free-energy constellation. Without conductivity a wall is therefore trapped in a potential minimum; with conductivity, a deeper, more favorable, minimum can be created by shifting of magnetic moments. This is a kind of aging effect observed in metals, for example, as a diffusion of carbon or nitrogen to alternate interstitial sites.[48] In our case, since electrons are being shifted, the diffusion is much more rapid; at room temperature the wall buries itself in about 10^{-9} sec. As the temperature is lowered, the electron exchange is slowed and the domain wall needs more and more time to deepen its potential well.

Application of an oscillatory field causes a wall to move up the sides of its well; the electron atmosphere tends to follow this swing by a redistribution of the Fe^{2+}-Fe^{3+} arrangement. If the period of oscillation is long in comparison to the electron diffusion time, the wall behaves as if oscillating in a shallow well not deepened by electron interchange. This seems to be the situation at room temperature, where the wall resonance appears unimpeded by electron drag. Lowering the temperature and thus lengthening the relaxation time of the electron atmosphere, causes the wall resonance to degenerate into a relaxation spectrum controlled by electron-spin redistribution. At very low temperatures, when the electrons are essentially immobilized in the room-temperature-resonance region, a wall resonance should reappear at a higher frequency, since the wall oscillation now corresponds to a motion in the deepened well.

Two other considerations support the interpretation that the low-frequency loss peak is due to domain-wall motion rather than to domain rotation. Effective internal fields of the order of 240 to 2400 amp/m (3 to 30 oersteds) are required to account for gyromagnetic resonances in the range of 10^7 to 10^8 cps. Although such low fields might result from a near balance of shape and anisotropy effects, such a balance is improbable. Furthermore, the Landau-Lifshitz dispersion formula predicts that for increased damping, a gyromagnetic resonance should tend towards a relaxation spectrum at frequencies much greater than those of the original position; this is in contrast to our observations.

Origin of the Magnetic Dispersion between 10^8 and 10^{11} cps

The dispersion and loss peak resolved between 10^9 and 10^{10} cps in the room-temperature spectra of nickel-

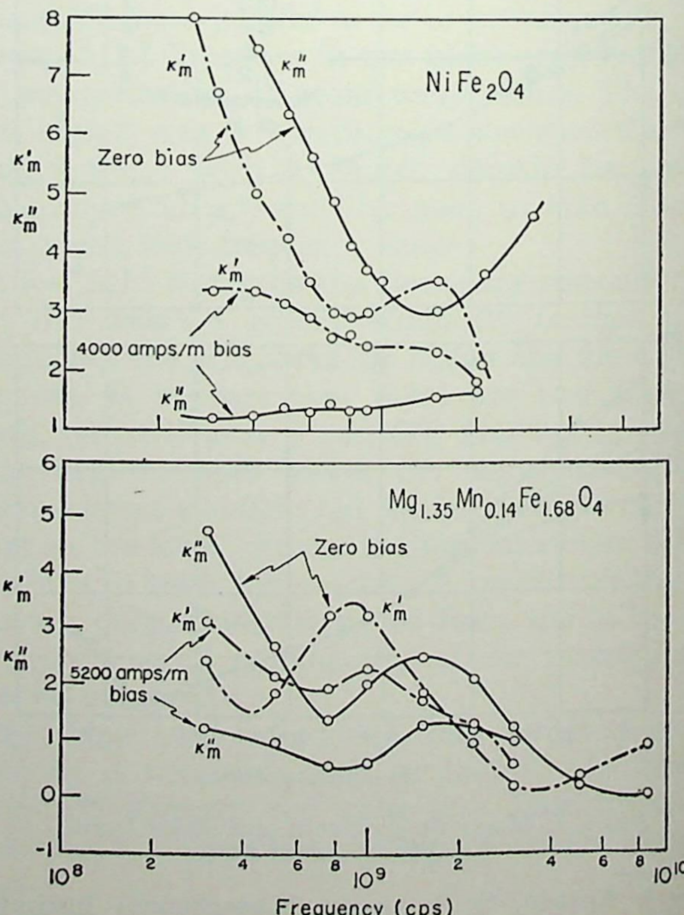


Fig. 7.10. Effect of static magnetic field on frequency response of nickel ferrites.

[48] J. L. Snoek, Physica 8, 711 (1941); L. Néel, J. phys. radium 13, 249 (1952); A. M. Clogston, Bell System Tech. J. 34, 739 (1955).

zinc ferrites is clearly a resonance, and in nickel ferrite the high frequency limit is in good agreement with that predicted for a gyromagnetic resonance in the presence of domain walls (Sec. 6b). This appears therefore to be the correct interpretation. The origin of the loss peak between 10^8 and 10^9 cps which develops out of the low-frequency dispersion as the temperature is lowered (Fig. 7.5) is less clear. It could be the re-emerging wall resonance referred to in the previous section, but more probably it is a group of gyromagnetic resonance produced when excitation takes place perpendicular to domain walls. In the simple case treated in Sec. 6b, excitation perpendicular or parallel to such walls gives two resonances separated by a frequency interval of the order of $\gamma\mu_0 M/2$ (5×10^9 cps for nickel ferrite). The observed separations of the two peaks are somewhat smaller, but this might be accounted for by a more realistic domain configuration.

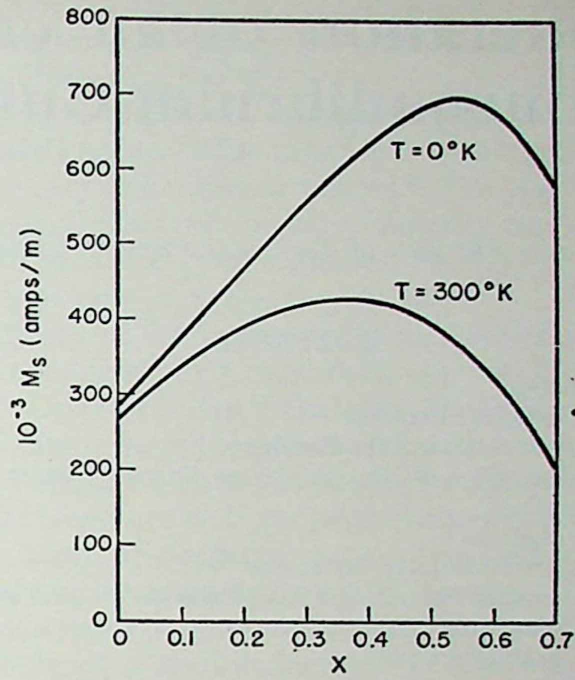


FIG. 7.12. Saturation magnetization for the nickel-zinc ferrite series.

Effect of Biasing Field

Application of a static magnetic field parallel to the oscillatory field markedly affects the lower resonance region (Fig. 7.10). The prominent minimum in κ_m' near 10^9 cps practically disappears, indicating that it was caused by the high-frequency tail of the lower dispersion. This behavior is consistent with the idea that this part of the dispersion is due to, or greatly influenced by, the presence of domain walls. In particular, the selective elimination of walls oriented perpendicularly to the oscillatory field would explain the reduction of the lower group of spin resonances whose position is determined by their presence.

Cation Substitution

The effect of cation substitution on magnetic spectra is not a simple one, since the fundamental properties of the material, such as saturation magnetization, Curie point, anisotropy, magnetostriction, conductivity, domain and grain structure, undergo changes. The temperature dependence of the initial permeability for the nickel-zinc ferrite series (Fig. 7.11) reflects relatively clearly that the Curie temperature decreases with increasing zinc content and that anisotropy and magnetostriction decrease as the Curie point is approached. The saturation magnetization, on the other hand, traverses a maximum and then falls to zero for zinc ferrite (Fig. 7.12).[49] The dispersion between 10^8 and 10^{10} cps, in a general way, adjusts to this change in magnetization by a shift in resonance frequency, but does not behave as simple theory would predict. This is probably due to variations in the other parameters and their effect on domain patterns and interactions.

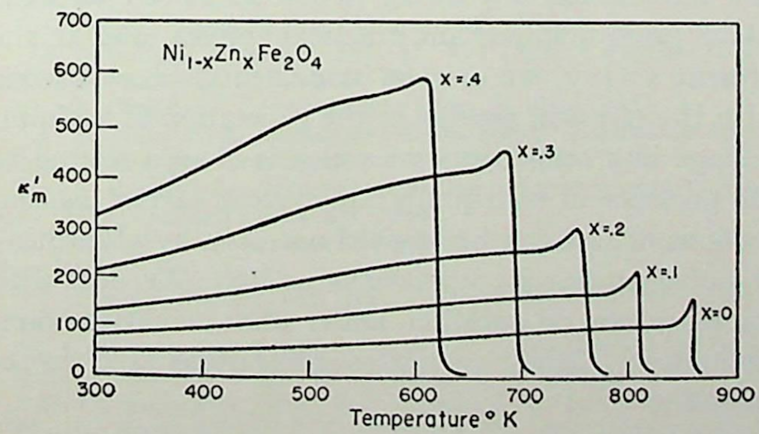


FIG. 7.11. Temperature dependence of initial permeability for nickel-zinc ferrites (10^3 cps).

It is obvious that we are still far from a complete understanding of these ferromagnetics. Many pieces of the puzzle, however, have been identified, and dielectric spectroscopy is one approach of assembling them in proper perspective.

ACKNOWLEDGMENTS

We wish to thank our co-workers, Professor D. J. Epstein and Dr. R. D. Waldron, whose contributions have been indicated in the text, for helpful discussions, and Mr. J. Kalnajs for his investigation of the optical properties of magnetite films.

[49] E. W. Gorter, Nature **165**, 798 (1950); C. Guillaud, J. phys. radium **12**, 143 (1951); R. Pauthenet, Ann. phys. [12] **7**, 714 (1952).

Generation of an emf in Semiconductors with Nonequilibrium Current Carrier Concentrations

JAN TAUC

Institute of Technical Physics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

CONTENTS

LIST OF SYMBOLS

b	ratio of electron to hole mobility
c	velocity of light
D_k	diffusion constants ($=\mathrm{k}T\mu_k/e$)
ζ'	Fermi level for thermal equilibrium (Fig. 3)
ζ_k'	chemical potentials of the assembly of electrons in conduction ($k=1$) or valence band ($k=2$) (Fig. 3)
ζ_k	chemical potentials of the assemblies of electrons and holes (Fig. 3)
$\Delta\zeta_k=\zeta_k-\zeta_{k0}$	change in chemical potential with illumination
e	absolute value of electron charge
E	electron energy
E_G	energy gap
ϵ_k	kinetic energy of an electron or a hole (Fig. 3)
$F=-\mathrm{grad}\,\varphi$	electric field
φ	potential of the electric field
g_f	number of electron-hole pairs produced in a unit volume per unit time
h	Planck's constant
I	electric current density
I_k	density of the electric current transferred by electrons or holes
I_{qk}	density of heat current transferred by electrons or holes
k	Boltzmann constant
k	index
$L_k=(D_k\tau)^{\frac{1}{2}}$	diffusion length of electrons or holes
m	free electron mass
m_k	effective mass of an electron or of a hole
μ_k	electron or hole mobility
$n_k=n_{k0}+\Delta n$	concentration of electrons in the conduction band or of holes in the valence band
Δn	concentration of electrons (=concentration of holes) added to the equilibrium concentrations
n_i	electron concentration in an intrinsic semiconductor in the dark
N_D, N_A	concentrations of donors and acceptors
$N=N_D-N_A$	
$\omega(E), \omega_k(\epsilon_k)$	number of permitted quantum states per unit energy interval per unit volume of the crystal
$Q_k^{*\prime}$	mean heat energy transferred by an electron or a hole
Q_k^*	mean kinetic energy transferred by an electron or a hole
ρ	electric resistivity
σ	electric conductivity
σ_k	electric conductivity of electrons or holes
$t_k=\sigma_k/\sigma$	transfer numbers
T	absolute temperature (ΔT temperature interval)
τ	mean lifetime of an electron-hole pair
U	voltage or emf
x	distance
Index:	0 denotes thermal-equilibrium state
	$k=1$ refers to electrons in the conduction band
	$k=2$ refers to holes in the valence band
	i means intrinsic semiconductor

1. INTRODUCTION

THE problem of direct transformation of solar or nuclear energy into electrical energy by means of the photovoltaic effect in semiconductors has aroused new interest recently owing to new kinds of photocells which promise important practical applications. In the present survey, we discuss fundamental ideas dealing with the physical aspects of the generation of a photovoltage in a semiconductor which is always related to the presence of nonequilibrium current carrier concentrations. These can be created not only by absorption of radiation but in other ways as well. We deal with the more general question, under what conditions nonequilibrium current carrier concentrations give rise to an emf.

1.1 Historical Remarks

The photovoltaic effect was first observed by Adams and Day in 1876 in selenium.[1] Later, much attention was devoted to photovoltaic effects in cuprous oxide

[1] W. G. Adams and R. E. Day, Proc. Roy. Soc. (London) A25, 113 (1877).

(Cu_2O). Coblenz[2] and Dember[3] observed the generation of an emf between the illuminated and nonilluminated parts of some semiconducting crystals, in particular Cu_2O. This effect was then studied by A. F. Joffe and his collaborators experimentally and theoretically. In 1927, Grondahl[4] and later Lange[5] observed the generation of an emf at the contact between Cu_2O and Cu, the so-called barrier-layer photovoltaic effect, the essential features of which were explained by Schottky and his collaborators.[6] An account of the literature of this older work is given in a review by Grondahl[7] and in the books by Joffe,[8] Boutry,[9] and Lange.[10]

This work made clear that the photovoltaic effect is not caused by the pressure of photons on the electrons as was originally supposed by some workers, but it is closely related to the diffusion of the current carriers. This point of view was particularly stressed by Frenkel.[11] Of great importance for explanation of the photovoltaic effect was a paper by Landau and Lifshitz[12] and another by Davydov[13] in which it was shown that it can be caused by the excess concentration of minority current carriers.

In 1939, Mott[14] dealt with the question of how the nonequilibrium concentration of minority carriers produces an emf in a barrier-layer photocell.

In addition to Se and Cu_2O the photovoltaic effect was observed in many other semiconductors. The Ag_2S (Bernatzkii and Heikman[15]) and Tl_2S (Kolomietz[16]) photocells were constructed in 1937. Of fundamental importance was the discovery of the photovoltaic effect in a *p-n* junction situated within a semiconductor crystal. The first observation was made by Ohl[17,18] in crystals of silicon; later observations were reported by Simpson, Soole, Sosnowski, and Starkiewicz[19–21] in PbS, by Benzer[22] and by Becker and Fan[23] in Ge. In recent years this effect was observed in semiconducting compounds, e.g., InSb,[24–27] GaAs,[28,29] InAs,[30] InP,[31] AlSb,[32] and CdS.[33,34]

Practical applications were found first for the selenium and cuprous oxide photocells which are used as light intensity indicators or meters.[35] The *p-n* junction silicon photocells attain a high efficiency for the conversion of solar energy into electrical energy[36,37] and are an essential part of a nuclear battery.[37,38]

The theory of the foregoing effects was elaborated in considerable detail by a number of workers (Lashkarev *et al.*,[39,40] Lehovec,[41] Fan,[42] Gubanov,[43] and Moss[44]).

In all these effects the generation of a stationary emf is related either to the presence of a rectifying potential barrier at a contact or in the bulk of the semiconductor (barrier-layer photoeffects) or to the presence of a nonrectifying semiconductor-metal contact (Dember-effect). However, a photovoltage can be produced even in the bulk of a semiconductor.[45,46] This recently observed effect is taken as the point of departure for consideration here as it permits the essential features of the photovoltaic effect to be shown in a particularly simple and general manner.

This paper is chiefly devoted to photovoltaic effects. However, in the last part we deal with the generation of an emf in cases when nonequilibrium carrier concentration is produced by other means than by absorption of radiation.

1.2 Formulation of the Problem

Let us consider the simple semiconductor circuit shown in Fig. 1. The circuit is interrupted at the points

[2] W. W. Coblentz, Natl. Bur. Standards, Sci. Papers (1919–1921).
[3] H. Dember, Physik. Z. 32, 554 and 856 (1931); Naturwissenschaften 20, 758 (1932).
[4] L. O. Grondahl and P. H. Geiger, Trans. Am. Inst. Elec. Engrs. 46, 357 (1927).
[5] B. Lange, Physik. Z. 31, 139 and 964 (1930).
[6] W. Schottky, Z. tech. Phys. 11, 458 (1930); Physik. Z. 31, 913 (1930); Physik. Z. 32, 833 (1931).
[7] L. O. Grondahl, Revs. Modern Phys. 5, 162 (1933).
[8] A. F. Joffe, *Semi-conducteurs électroniques* (Actualités scientifiques et industrielles, Hermann and Cie, Paris, 1935).
[9] G. A. Boutry, *Les phénomènes photoélectriques et leurs applications* (Hermann and Cie, Paris, 1936).
[10] B. Lange, *Photoelements* (Reinhold Publishing Corporation, New York, 1938).
[11] J. Frenkel, Nature 132, 312 (1933); Physik. Z. Sowjetunion 8, 185 (1935).
[12] L. Landau and E. Lifshitz, Phys. Z. Sowjetunion 9, 477 (1936).
[13] V. I. Davydov, Tech. Phys. USSR 5, 79 (1938).
[14] N. F. Mott, Proc. Roy. Soc. (London) A171, 281 (1939).
[15] V. K. Bernackii and D. S. Heikman, Fiz. Zap. Ukrainskoi Ak. nauk, Kiev, 7, 69 (1938).
[16] B. T. Kolomietz, Doklady Akad. Nauk SSSR 19, 383 (1938).
[17] R. S. Ohl, U. S. Patent 2,402,662, filed May 27, 1941.
[18] J. H. Scaff and H. C. Theuerer, Trans. Am. Inst. Mining Met. Engrs. 185, 383 (1949).
[19] Starkiewicz, Sosnowski, and Simpson, Nature 158, 28 (1946).
[20] L. Sosnowski, Phys. Rev. 72, 641 (1947).
[21] Sosnowski, Soole, and Starkiewicz, Nature 160, 471 (1947).
[22] S. Benzer, Phys. Rev. 72, 1267 (1947).
[23] M. Becker and H. Y. Fan, Phys. Rev. 78, 301 (1950).
[24] J. Tauc and A. Abrahám, Czechosl. J. Phys. 4, 478 (1954).
[25] Avery, Goodwin, Lawson, and Moss, Proc. Phys. Soc. (London) B67, 761 (1954).
[26] H. P. R. Frederikse and R. F. Blunt, Proc. Inst. Radio Engrs. 43, 1828 (1955).
[27] Kurnick, Goldberg, Mitchell, and Zitter, Abstract No. 98, Spring Meeting Electrochemical Society Cincinnati, Ohio (May, 1955).
[28] R. Gremmelmaier, Z. Naturforsch. 10a, 501 (1955).
[29] H. Pfister, Z. Naturforsch. 11a, 434 (1956).
[30] R. M. Talley and D. P. Enright, Phys. Rev. 95, 1092 (1954).
[31] R. Gremmelmaier and H. Welker, Z. Naturforsch. 11a, 420 (1956).
[32] A. Abrahám, Czechosl. J. Phys. 6, 624 (1956).
[33] Reynolds, Leies, Antes, and Marburger, Phys. Rev. 96, 533 (1954).
[34] D. C. Reynolds and S. J. Czyzak, Phys. Rev. 96, 1705 (1954).
[35] V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application* (John Wiley and Sons, Inc., New York, 1950).
[36] Chapin, Fuller, and Pearson, J. Appl. Phys. 25, 676 (1954).
[37] W. G. Pfann and W. van Roosbroeck, J. Appl. Phys. 25, 1422 (1954).
[38] P. Rappaport, Phys. Rev. 93, 246 (1954).
[39] V. E. Lashkarev and K. M. Košonogova, Zhur. Eksptl. i Teort. Fiz. 16, 786 (1946); *ibid.* 18, 927 (1948).
[40] V. E. Lashkarev, Zhur. Eksptl. i Teort. Fiz. 18, 917 (1948); Izvest Akad. Nauk SSSR. Ser. Fiz. 16, 18 (1952).
[41] K. Lehovec, Z. Naturforsch. 2, 398 (1947); Phys. Rev. 74, 463 (1948).
[42] H. Y. Fan, Phys. Rev. 75, 1631 (1949).
[43] A. J. Gubanov, Zhur. Eksptl. i Teort. Fiz. 25, 307 (1953).
[44] T. S. Moss, J. Electronics 1, 126 (1955).
[45] J. Tauc, Czechosl. J. Phys. 5, 178 (1955).
[46] Z. Trousil, Czechosl. J. Phys. 6, 96 (1956).

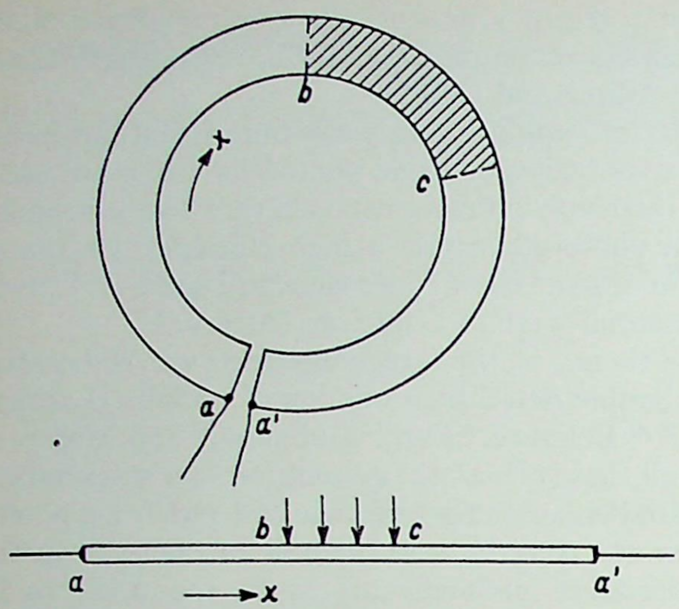


FIG. 1. Semiconductor circuit. Part b, c is illuminated.

a and a' where the leads for measuring the emf by the potentiometric method are connected. In part b, c there is a nonequilibrium current carrier concentration produced, e.g., by illumination. The contacts a and a' are sufficiently distant from the illuminated part so that in their neighborhood the nonequilibrium concentrations of current carriers are zero.

The question is what emf is measured between the contacts a and a' in the stationary case? A necessary condition for generation of an emf is some asymmetry of the circuit. What is the nature of this asymmetry?

Effects at the contacts semiconductor-metal, which we shall not consider in the present paper, are eliminated in the circuit shown in Fig. 1. The boundary effects here take place in the neighborhood of the points b and c situated within the semiconductor of known parameters instead at the contacts of a semiconductor with a metal, the properties of which have not yet been thoroughly investigated. The leads are connected at the points a and a' of identical chemical composition, at the same temperature, situated in the same magnetic field, etc. We thus respect a theorem[47,48] by Gibbs that electrical potential difference is uniquely defined only between two bodies of the same chemical composition and at the same temperature. This circumstance is important since from the history of the photovoltaic effect it is well known that some investigators did not respect it, and thus obtained erroneous results in the case of extrinsic semiconductors, in which illumination produces current carriers of the same type as those present in the dark. Landau and Lifshitz pointed out[12] that the voltage generated within the semiconductor in this case is compensated by the change of the contact potential if the current carrier energy distribution is governed by the Maxwell-Boltzmann statistics.[49]

We simplify the problem by the following assumptions. (α) We deal with semiconductors of the same type as germanium, that is, with large current carrier mobilities, of comparatively high electrical conductivity, practically without traps. Almost all of the whole crystal is therefore practically electrically neutral. The neutrality condition cannot hold exactly since the internal electric field in the crystal can be produced only by the presence of electric charges; the meaning of the neutrality condition is given in Sec. 2.5. It is, of course, possible that considerable space charge densities are located in regions of abrupt inhomogeneities (e.g., near the surface or at a p-n junction). Nevertheless, most of of the volume of the crystal is nearly electrically neutral and keeps its neutrality even if illuminated. This is not the case, for example, in ionic crystals in which space charges often accompany photoelectric effects. The assumptions mentioned are well satisfied in semiconductors with prevailing covalent binding, such as Ge, Si, $A^{III}B^{V}$ and other intermetallic semiconducting compounds.

(β) For the sake of simplicity we assume the geometry to be linear. This means that the thickness of the sample is small compared with the reciprocal value of the absorption constant or with the diffusion length of the carriers so the concentrations are practically constant over the whole cross section of the sample.

1.3 Generation of Nonequilibrium Current Carrier Concentration

In a semiconductor nonequilibrium concentrations n_1 and n_2 of electrons and holes can exist differing from the concentrations, which correspond to the thermal equilibrium n_{10}, n_{20}, by amounts $\Delta n_1 = n_1 - n_{10}$ and

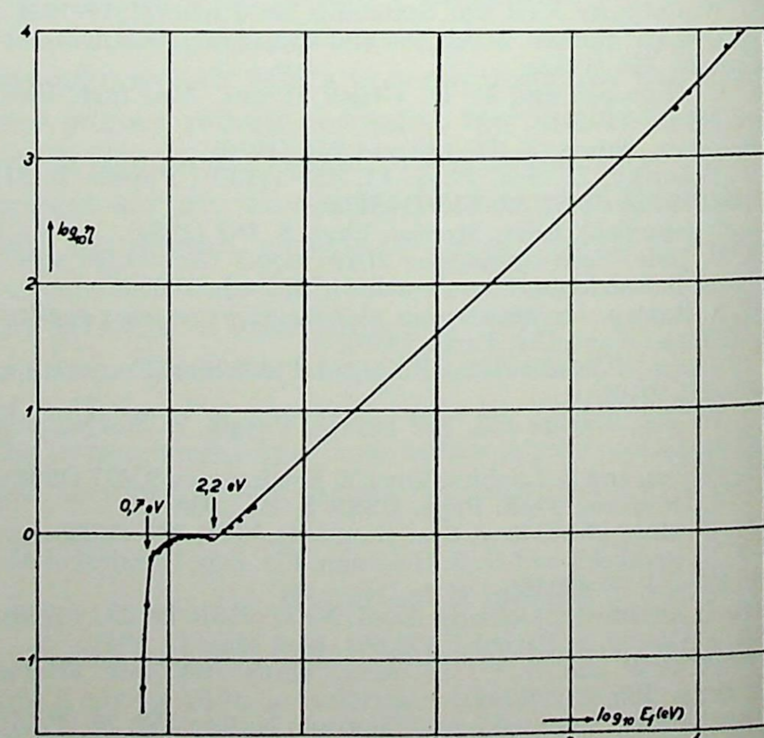


FIG. 2. Dependence of quantum efficiency η of the inner photoelectric effect[50,51] in germanium on the photon energy E_f.

[47] E. A. Guggenheim, J. Phys. Chem. 33, 842 (1929).
[48] E. A. Guggenheim, *Thermodynamics* (North-Holland Publishing Company, Amsterdam, 1950).
[49] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, New York, 1940), p. 192.

$\Delta n_2 = n_2 - n_{20}$. If there are no traps in the semiconductor and the condition of electrical neutrality is fulfilled [this is the case if assumption (α) holds] one has practically $\Delta n_1 = \Delta n_2 = \Delta n$. The value of nonequilibrium concentrations is determined by two processes: the rate of generation of excess current carriers, and their rate of recombination.

An important way of generating added carriers is by absorption of radiation. If $n_f dv$ photons or particles are absorbed in a volume dv per second from which each produces η electron-hole pairs with a mean lifetime τ, an excess concentration Δn is created given by $\Delta n = \eta n_f \tau = g_f \tau$, neglecting the diffusion of carriers from or into the volume considered.

The particularly important number η is denoted as the quantum efficiency of the inner photoelectric effect. Its dependence[50,51] on the photon energy E_f was measured in germanium (Fig. 2). A photon is able to produce an electron-hole pair only in case its energy E_f exceeds some minimum energy which corresponds to the infrared absorption limit of the semiconductor. In semiconductors satisfying assumptions (α) and [(ϵ) (see Sec. 3.1)] it is given (at least approximately) by the value of the energy gap E_G. In the neighborhood of the absorption limit towards shorter wavelengths ($E_f > E_G$), we have in germanium $\eta \approx 1$ until we reach an energy $\epsilon' = 2.2$ ev. For $E_f > \epsilon'$ the quantum efficiency is proportional to E_f; the energy needed to produce an electron-hole pair is constant and equal to $\epsilon = 2.5$ ev. In this energy range the kinetic energy of electrons released by the absorption of a high-energy photon is used for generation of additional electrons and holes. The energy $\epsilon - E_G$ is probably transferred to the lattice vibrations. The energy ϵ has been measured[52] with high-energy radiation in some other semiconductors.

Another way of producing nonequilibrium concentrations is diffusion of current carriers from regions of higher concentrations into regions of lower concentrations. The regions of higher concentrations are, e.g., irradiated regions, regions with a higher impurity content, or at a higher temperature.

2. BASIC EQUATIONS

We start from the well-known equations generally used for description of electronic effects in semiconductors: the equations for the current densities, the continuity equation and the Poisson equation. In writing the current density equations we use notation familiar[53] in the thermodynamics of irreversible processes. This helps to show the analogy between generation of an emf in semiconductors and general laws governing energy transformations, e.g., in a galvanic

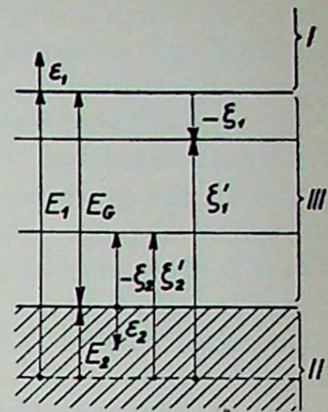


FIG. 3. Notation of energies: *I* conduction band, *II* valence band, and III energy gap.

cell. In deriving the equations for current densities it is usual to start from an equation such as Boltzmann's equation for the stationary state and to consider, in final formulas, certain expressions as phenomenological constants which are to be determined experimentally.

We take here a somewhat different point of view. We regard the equations of current densities as purely phenomenological equations which are valid generally, without any assumptions about the zone structure of the semiconductor or the electron scattering mechanism. Of course, their generality is understood in the same sense as the generality of the equations of the thermodynamics of irreversible processes, that is, as a first approximation of more exact kinetic equations which can be developed into approximations of higher orders (compare Sec. 3 of DeGroot[53]).

2.1 Current Density Equation

These can be written in the form (DeGroot,[53] Sec. 52):

$$I_k = \frac{1}{e}\sigma_k\left(-e\,\mathrm{grad}\varphi + T\,\mathrm{grad}\left(\frac{\zeta_k'}{T}\right) + Q_k^{*\prime}\frac{1}{T}\,\mathrm{grad}T\right). \quad (1)$$

The index $k=1$ refers here to electrons in the conduction band, $k=2$ to electrons in the valence band. The energy levels are shown in Fig. 3. This holds for stationary states provided two assumptions are valid: the current carriers undergo many collisions in time intervals which are considered in the theory and further that the change in the electrostatic energy of an electron over a distance equal in order of magnitude to the mean free path of an electron is small compared with the mean thermal energy. The first assumption excludes electric fields of very high frequencies, which are not considered here owing to restriction to stationary states, and further requires that the mean lifetime of an electron-hole pair is sufficiently longer than the relaxation time. The second assumption excludes very high electric fields such as are present, e.g., in a p-n junction. Inside such localized potential barriers the energy distribution of current carriers is substantially different from the dis-

[50] S. Koc, Czechosl. J. Phys. 7, 91 (1957).
[51] Drahokoupil, Malkovská, and Tauc, Czechosl. J. Phys. 7, 57 (1957).
[52] K. G. MacKay, Phys. Rev. 84, 829 (1951).
[53] S. R. DeGroot, *Thermodynamics of Irreversible Processes* (North-Holland Publishing Company, Amsterdam; Interscience Publishers, Inc., New York, 1952).

tribution which corresponds to the temperature of the lattice T, and other ways are necessary to calculate the current densities (Sec. 5).

Equation (1) expresses the fact that, on the assumptions just mentioned, the current densities are proportional to the gradients of the electric potential φ, of the chemical potentials ζ_k' and of the temperature T. There is theoretical and experimental evidence[54,55] that if regions of localized potential barriers are excluded even excess current carriers created by photon absorption practically have the energy distribution corresponding to T. This is owing to the fact that interaction of carriers with the lattice is so strong that after their liberation by light they lose[56] their excess energy very quickly (say within 10^{-11} sec). It is thus possible to define the chemical potentials of the assembly of electrons in the conduction and the valence band and other quantities for the case that current carrier concentrations do not correspond to thermal equilibrium.

2.2 Chemical Potentials of the Assemblies of Electrons and Holes

From the Fermi-Dirac statistics the following relation holds between the concentration of electrons in the conduction band n_1 and their chemical potential ζ_1',

$$n_1=\int_{E_1}^{\infty}\frac{\omega(E)dE}{1+e^{(E-\zeta_1')/kT}}. \tag{2}$$

Similarly, for the concentration of holes in the valence band,

$$n_2=\int_{-\infty}^{E_2}\omega(E)\left(1-\frac{1}{1+e^{(E-\zeta_2')kT}}\right)dE. \tag{3}$$

Here $\omega(E)dE$ denotes the number of permitted energy states of an electron in the interval E, $E+dE$. For the case that the assembly of electrons in the conduction band is in thermal equilibrium with the assembly of electrons in the valence band,

$$\zeta_1'=\zeta_2', \tag{4}$$

and the Fermi level is defined. If this is not the case chemical potentials of both assemblies are different; they were introduced by Shockley[57–59] into the theory of semiconductors and denoted as "quasi-Fermi levels."

It is sometimes advantageous to introduce the following notation (Fig. 3):

in the conduction band: $E=E_1+\epsilon_1$, $\zeta_1'=E_1+\zeta_1$,

in the valence band: $E=E_2-\epsilon_2$, $\zeta_2'=E_2-\zeta_2$.

[54] A. B. Giordano *et al.*, Phys. Rev. 88, 1368 (1952).
[55] S. M. Ryvkin, Zhur. Tech. Fiz. 24, 2136 (1954).
[56] B. I. Davydov, Zhur. Eksptl. i. Teort. Fiz. 7, 1069 (1937); *ibid.* 2212 (1937).
[57] W. Shockley, Bell System Tech. J. 28, 435 (1949).
[58] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950).
[59] Shockley, Sparks, and Teal, Phys. Rev. 83, 151 (1951).

Equations (2) and (3) then take the simple form

$$n_k=\int_0^{\infty}\frac{\omega_k(\epsilon_k)d\epsilon_k}{1+e^{(\epsilon_k-\zeta_k)/kT}}. \tag{5}$$

The quantities ζ_k will be referred to as chemical potentials of the assembly of electrons (ζ_1) and of the assembly of holes (ζ_2). Using them the thermal equilibrium condition (4) is written

$$\zeta_{10}+\zeta_{20}=-E_G. \tag{6}$$

The density of energy states $\omega_k(\epsilon_k)$ depends on the Brillouin-zone structure. In the simple case of spherical energy surfaces we have

$$\omega_k(\epsilon_k)=4\pi\left(\frac{2m_{dk}}{h^2}\right)^{\frac{3}{2}}\epsilon_k^{\frac{1}{2}}, \tag{7}$$

where $m_{dk}=m_k$ and m_k is the effective mass of an electron ($k=1$) or of a hole ($k=2$). In the case of elliptical energy surfaces where the effective masses are different in the three fundamental directions m_k', m_k'', m_k''', (7) holds with $m_{dk}=(m_k'm_k''m_k''')^{\frac{1}{3}}$. If each surface of constant energy is a set of N_s similar ellipsoids centered on the cubic axes of the Brillouin zone (which is probably the case for the conduction band of germanium with $N_s=4$ or 8) one has $m_{dk}=(N_s^2m_k'm_k''m_k''')^{\frac{1}{3}}$.

Determination of chemical potentials is especially simple for the case that classical statistics hold, that is, if $-\zeta_k/kT\gg1$. Then

$$n_k=e^{\zeta_k/kT}\int_0^{\infty}\omega_k(\epsilon_k)e^{-\epsilon_k/kT}d\epsilon_k=A_k(T)e^{\zeta_k/kT}, \tag{8}$$

where the partition function $A_k(T)$, for the case that the effective mass concept can be used, has the form

$$A_k(T)=2\left(\frac{2\pi m_{dk}kT}{h^2}\right)^{\frac{3}{2}}. \tag{9}$$

Introducing the chemical potentials ζ_k instead of ζ_k' into Eq. (1) and supposing that E_1, E_2 do not depend on x (which can be justified particularly in the case that E_G is independent of x), we obtain

$$I_1=\frac{\sigma_1}{e}\left(-e\,\mathrm{grad}\varphi+T\,\mathrm{grad}\left(\frac{\zeta_1}{T}\right)+Q_1^*\frac{1}{T}\mathrm{grad}T\right)$$
$$=\frac{\sigma_1}{e}\left(-e\,\mathrm{grad}\varphi+\mathrm{grad}\zeta_1+S_1^*\,\mathrm{grad}T\right), \tag{10}$$

$$I_2=\frac{\sigma_2}{e}\left(-e\,\mathrm{grad}\varphi-T\,\mathrm{grad}\left(\frac{\zeta_2}{T}\right)-Q_2^*\frac{1}{T}\mathrm{grad}T\right)$$
$$=\frac{\sigma_2}{e}\left(-e\,\mathrm{grad}\varphi-\mathrm{grad}\zeta_2-S_2^*\,\mathrm{grad}T\right). \tag{11}$$

The new factors are related to the old ones as follows:

$$Q_1^*=Q_1^{*\prime}-E_1, \tag{12}$$

$$Q_2^*=E_2-Q_2^{*\prime}, \tag{13}$$

$$S_k^*=(1/T)(Q_k^*-\zeta_k). \tag{14}$$

2.3 Quantities of Transfer $Q_k^{*\prime}$, Q_k^*, S_k^*

The physical significance of the transfer quantities $Q_k^{*\prime}$ is best seen if, together with Eq. (1), we consider the equation for the heat current density transported by electrons in the conduction or valence band, which according to the Onsager principle,[60] can be written

$$I_{qk}=-\frac{1}{e^2}\sigma_k\left\{Q_k^{*\prime}\left[-e\,\mathrm{grad}\varphi+T\,\mathrm{grad}\left(\frac{\zeta_k'}{T}\right)\right]+P_k^{*\prime}\frac{1}{T}\,\mathrm{grad}T\right\} \tag{15}$$

($P_k^{*\prime}$ are constants). If $\mathrm{grad}T=0$, we have

$$Q_k^{*\prime}=I_{qk}/I_k/(-e). \tag{16}$$

$Q_k^{*\prime}$ is the heat current density in the conduction ($k=1$) or the valence ($k=2$) band which corresponds to unit current density of electrons (one electron per 1 cm^2 per sec) for the case that $\mathrm{grad}T=0$. We call $Q_k^{*\prime}$ the heat transfer due to electrons in the conduction or valence band.

These quantities are determined by kinetic considerations. We speak here explicitly about the conduction band. The heat transfer $Q_1^{*\prime}$ is composed of two parts, the electronic part $Q_{1e}^{*\prime}$ and the phonon-drag part $Q_{1p}^{*\prime}$. The electronic part $Q_{1e}^{*\prime}$ can be determined as follows. For the electric current density we have for the case of spherical zone structure (supposing the validity of some assumptions[61] about electron scattering probabilities)

$$I_1=-\frac{16\pi m_{d1}e^2}{3h^3}\int_0^\infty l_1(\epsilon_1)\epsilon_1\frac{\partial f_0}{\partial\epsilon_1}d\epsilon_1, \tag{17}$$

where $l_1(\epsilon_1)$ is the mean free path of electrons and f_0 is the Fermi distribution function

$$f_0=\frac{1}{1+e^{(\epsilon_1-\zeta_1)/kT}}. \tag{18}$$

Every electron which takes part in the transport of the electric current I_1 transfers an energy $E=E_1+\epsilon_1$; therefore for I_{q1e}, we obtain

$$I_{q1e}=\frac{16\pi m_{d1}e}{3h^3}\int_0^\infty l_1(\epsilon_1)\epsilon_1(E_1+\epsilon_1)\frac{\partial f_0}{\partial\epsilon_1}d\epsilon_1$$

$$=-\frac{1}{e}(E_1+Q_{1e}^*)I_1, \tag{19}$$

where Q_{1e}^* is the mean kinetic electron energy of transfer given by the following relation (with $k=1$):

$$Q_{ke}^*=\frac{\int_0^\infty l_k(\epsilon_k)\epsilon_k^2(\partial f_0/\partial\epsilon_k)d\epsilon_k}{\int_0^\infty l_k(\epsilon_k)\epsilon_k(\partial f_0/\partial\epsilon_k)d\epsilon_k}. \tag{20}$$

The same expression is valid for holes (with $k=2$). As was pointed out by Price,[62] Eq. (20) is correct only when the energy change on scattering is small compared with kT. If classical statistics are valid and energy zones are spherical, one has

for thermal scattering $l=\mathrm{const}$, $Q_{ke}^*=2kT$, (21)

for ionized impurities scattering $l\sim\epsilon^2$, $Q_{ke}^*=4kT$. (22)

The intermediate cases are treated by Anselm and Klachkin.[63]

The heat transfer $Q_{1p}^{*\prime}=Q_{1p}^*+E_1$ is caused by drag of low-frequency phonons on the electrons in the conduction band. This effect was first discussed by Gurevich[64] and later by Klemens[65]; it was experimentally discovered by Frederikse[66] and Geballe[67] on measuring the thermoelectric power of germanium at low temperatures. The theory of the effect was given by Frederikse[68] and by Herring.[69]

The value of Q_{ke}^* at room temperature is about 0.05–0.1 ev. The value of Q_{kp}^* can attain hundredths of electron volts at room temperature but increases considerably with decreasing temperature; in p-type silicon values as high as 18 ev have been observed.[70]

The quantities S_k^* are the entropies transferred by an electron or a hole.

2.4 Continuity Equation

When assumption (α) (Sec. 1.2) holds, the continuity equations for the current of electrons and holes can be written for the stationary case ($\partial n_k/\partial t=0$) as follows:

$$\mathrm{div}I_1=-\mathrm{div}I_2=eR-eg_f. \tag{23}$$

[60] Compare deGroot,[53] Chap. 1. A discussion of its application to semiconductors is given by P. J. Price [Phys. Rev. **104**, 1223 (1956)]. I am indebted to Dr. Price for sending me his manuscript prior to its publication.

[61] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Oxford University Press, New York, 1936).

[62] P. J. Price (private communication).

[63] A. I. Anselm and V. I. Klachkin, Zhur. Exptl. i Teort. Fiz. 22, 297 (1952).

[64] L. Gurevich, J. Phys. U.S.S.R. **9**, 477 (1945); *ibid.* **10**, 67 (1946).

[65] P. G. Klemens, Australian J. Phys. **7**, 520 (1954).

[66] H. P. R. Frederikse, Phys. Rev. **91**, 491 (1953).

[67] T. H. Geballe, Phys. Rev. **92**, 857 (1953).

[68] H. P. R. Frederikse, Phys. Rev. **92**, 248 (1953).

[69] C. Herring, Phys. Rev. **92**, 857 (1953); *ibid.* **96**, 1163 (1954).

[70] T. H. Geballe and G. W. Hull, Phys. Rev. **98**, 940 (1955).

Here R denotes the recombination function (the number of excess electrons and holes recombining in 1 cm³ per sec). In the most simple case[71] of linear recombination, valid for small concentration of excess current carriers, $\Delta n \ll n_{10}$ or n_{20} and $R=\Delta n/\tau$.

2.5 Poisson Equation

The Poisson equation has the following form:

$$\text{div}F=(e/K)(n_2-n_1+N_D{}^+-N_A{}^-)=(e/K)\delta n. \quad (24)$$

Here $N_D{}^+$, $N_A{}^-$ are concentrations of ionized donors and acceptors; $e\delta n$ is the space-charge density, K the dielectric constant. We introduce now the following assumption (γ), which holds approximately for semiconductors of the type defined by assumption (α) and for not too low temperatures. (γ) The donors and acceptors are fully ionized whether carrier concentrations correspond to thermal equilibrium or not. We then have $N_D{}^+=N_D-N_A$ in n-type semiconductors, $N_A{}^-=N_A-N_D$ in p-type semiconductors.

Assumption (γ) means a great simplification but can be used only for certain kinds of semiconductors, certain kinds of impurities and in a certain temperature range. For the space-charge density $e\delta n$ we consider two limiting cases.

(a) *Electrical Neutrality*

Small disturbances of electrical neutrality give rise to strong electric fields so that strong forces are always present which tend to restore neutrality. Of course, it is not possible to have complete electrical neutrality because electric fields producing the emf have to be present. But if these fields are feeble, and their variation along the sample is continuous, δn is quite neglegible compared with n_1 and n_2. For instance, an electric field having a gradient 100 v/cm corresponds in germanium ($K_{\text{rel}}=16$) to a quite negligible concentration $\delta n=K/e\,\text{grad}F=7.5\cdot10^8$ cm^{-3}. It is possible then in every point in the semiconductor (with the exception of localized potential barriers) to put into the equation for current densities and into the continuity equation the condition of electrical neutrality

$$n_1-n_{10}=n_2-n_{20}=\Delta n. \quad (25)$$

(b) *Strong Space Charge*

Strong space-charge density is present in regions of large changes of some semiconductor parameters, e.g., near the surface, in a p-n junction, etc. There are strong electric fields inside such regions and Eq. (1) is not valid there (Sec. 5). Under the conditions existing in a p-n junction discussed by Shockley[57] the space-charge density of current carriers can be neglected and the Poisson equation is reduced to the simple form

$$\text{div}F=(e/K)(N_D-N_A). \quad (26)$$

[71] W. Shockley and W. T. Read, Jr., Phys. Rev. **87**, 835 (1952).

From this equation, the potential distribution φ in a p-n junction can be determined (Sec. 5.1).

3. GENERATION OF AN emf

Let us return to the case illustrated in Fig. 1. We now discuss the theoretical expression for the emf generated in the circuit on the assumption that in the whole circuit the current carriers are in thermal equilibrium with the lattice; Eqs. (1), (23), and (25) are then valid at every point. Such a condition is realized for the case that the gradients of the parameters of the semiconductor (e.g., such as the impurity concentration) in the sample are small. In Sec. 5 we consider the case that inside the illuminated parts b and c there is a localized p-n junction in which Eq. (1) does not hold. It turns out that essential conclusions reached by consideration of the quasi-equilibrium case can be extended to the other case.

In the one-dimensional geometry [assumption (β)], we get from (1) for the electric field $F=-\text{grad}\varphi$,

$$F=\frac{I}{\sigma}-\frac{1}{e}t_1T\frac{d}{dx}\left(\frac{\zeta_1'}{T}\right)-\frac{1}{e}t_2T\frac{d}{dx}\left(\frac{\zeta_2'}{T}\right)$$
$$-\frac{1}{e}t_1Q_1{}^{*\prime}\frac{1}{T}\frac{dT}{dx}-\frac{1}{e}t_2Q_2{}^{*\prime}\frac{1}{T}\frac{dT}{dx}, \quad (27)$$

on introducing the transfer numbers $t_k=\sigma_k/\sigma$.

To obtain the expression for the emf which can be measured between the contacts a and a' we must put $I=0$ and integrate (27) along the whole circuit of Fig. 1, obtaining

$$U=\oint Fdx=U_b+U_t, \quad (28)$$

$$U_b=-\frac{1}{e}\oint T\sum_{k=1,2}t_kd\left(\frac{\zeta_k'}{T}\right), \quad (29)$$

$$U_t=-\frac{1}{e}\oint\frac{1}{T}\sum_{k=1,2}t_kQ_k{}^{*\prime}dT. \quad (30)$$

The sign of U is chosen so that $U>0$ if the contact a is positive.

These expressions are simplified for the case that $dE_G/dx=0$. From (29) and (30), we get

$$U_b=-\frac{1}{e}\oint(t_1d\zeta_1-t_2d\zeta_2)$$
$$=-\frac{1}{e}\oint(t_1d\Delta\zeta_1-t_2d\Delta\zeta_2), \quad (31)$$

$$U_t=-\frac{1}{e}\oint(t_1S_1{}^*-t_2S_2{}^*)dT, \quad (32)$$

in which $\Delta\zeta_k=\zeta_k-\zeta_{k0}$.

3.1 Conditions for Generation of an emf

Conditions for generation of an emf can be obtained by discussion of the integrals (29) to (32).

Case (a) $dE_G/dx=0,\ dT/dx=0$

In this case $U_t=0$. From (31) we see that $U_b=0$ if there are no nonequilibrium carrier concentrations so that $\Delta\zeta_k=0$. Similarly, $U_b=0$ for the case that the semiconductor in the dark is homogeneous so that $d\zeta_{10}=-d\zeta_{20}$ and $b=\mu_1/\mu_2=\text{const}$, because then t_k and $\Delta\zeta_k$ are functions of Δn only and integration over the closed circuit in (31) gives zero. Further, the integral (31) is always equal to zero if we have only one kind of current carriers, that is if $t_1=1,\ t_2=0$ or $t_1=0,\ t_2=1$.

The condition for generation of an emf is the simultaneous presence of nonequilibrium carrier concentration and of an inhomogeneity as the integral (31) differs from zero only if t_k and $\Delta\zeta_k$ vary along the sample so that they cannot be expressed as a single-valued function of one variable. Consequently, the presence of two independent factors determining t_k and $\Delta\zeta_k$ is necessary.

The most important inhomogeneity is the variation of the concentration of donors or acceptors along the sample (Sec. 4). Another possibility is the variation of the ratio of the current carrier mobilities $b=\mu_1/\mu_2$* or of the effective masses of current carriers along the sample. The effects are usually very small. A further possibility is the variation of magnetic field intensity along the sample. In this case, (Sec. 6.1), generation of an emf is due to the dependence of the chemical potentials ζ_k on the magnetic field. On the other hand, the variation of recombination center concentrations along an otherwise homogeneous sample (which influence only the mean lifetime of current carriers) cannot give rise to an emf as it causes the variation only of Δn, equivalent to a nonhomogeneous illumination of a homogeneous sample.

Case (b) $dE_G/dx=0,\ dT/dx\neq 0$

For the case that a temperature gradient is present in the sample an emf is generated even if no nonequilibrium current carriers are present; (32) gives then the usual thermoelectric voltage. In calculating the absolute thermoelectric power of a semiconductor it is expedient to imagine a circuit composed of the semiconductor and of an electrical conductor in which it is possible to put $S^*=0$ (Fig. 4). The integration in (32) is easily carried out if $\Delta T\ll T$, and we get[72] the well-known formula for the thermoelectric power:

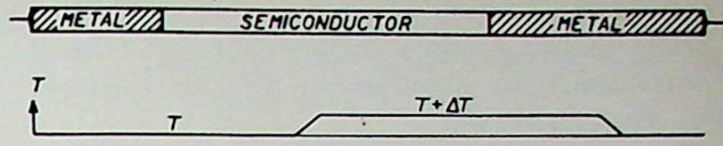


FIG. 4. Measurement of the thermoelectric power.

$$u=\frac{U_t}{\Delta T}=-\frac{1}{e}(t_{10}S_{10}^*-t_{20}S_{20}^*)$$
$$=\frac{k}{e}\left(t_{10}\frac{\zeta_{10}-Q_1^*}{kT}-t_{20}\frac{\zeta_{20}-Q_2^*}{kT}\right). \tag{33}$$

Presence of nonequilibrium carrier concentrations causes a change in thermoelectric voltage in the circuit. This we call a photothermoelectric effect (Sec. 6.2).

Case (c) $dE_G/dx\neq 0$

If E_G cannot be considered as constant in the sample an emf can be generated which is easily calculated from (29). If the sample is assumed to be otherwise homogeneous we get [if $\Delta\sigma=\Delta\sigma_s=\text{const}$ inside the part b, c, $\Delta\sigma=0$ outside, (Sec. 3.2, assumption (ζ))]:

$$U_b=-\frac{1}{e}\Delta t_1(E_{Gc}-E_{Gb}), \tag{34}$$

where $\Delta t_1=t_1-t_{10}$ and E_{Gc}, E_{Gb} are the values of the energy gap at the points c and b. This emf plays a role in the thermal photovoltaic effect owing to the dependence of E_G on the temperature (Sec. 6.2). It would also be of importance in a photovoltaic effect suggested by Price[63] where the inhomogeneity of the semiconductor is caused by strain.

Equations (29) to (32) hold for semiconductors even if assumptions (α) and (γ) are not valid. However, in performing the actual calculations in the following sections we have limited our considerations by the assumptions (α), (β), and (γ) and also the following two: (δ) The ratio of the electron-hole mobilities $b=\mu_1/\mu_2$ is constant throughout the sample. (ϵ) For the assemblies of electrons and holes classical statistics hold; that is, n_k is assumed to be always small compared with $A_k(T)$ even in the illuminated part of the semiconductor.

3.2 Distribution of Current Carrier Concentrations

From (29) to (32) one can calculate the emf generated in the circuit of Fig. 1, provided that the concentrations of current carriers are known in every point of the semiconductor from which ζ_k and t_k can easily be calcu-

* For example, if in an otherwise homogeneous semiconductor (n_{k0} are constant throughout the sample) the values of b at the points b and c are b_b and b_c the integral (31) gives for the emf:

$$U=U_b=\frac{k}{e}\left[\frac{b_c-1}{b_c+1}\frac{\Delta\sigma_{sc}}{\sigma_{0c}}-\frac{b_b-1}{b_b+1}\frac{\Delta\sigma_{sb}}{\sigma_{0b}}\right],$$

where σ_{0c}, σ_{0b} are dark conductivities at the points c and b; $\Delta\sigma_{sc}$, $\Delta\sigma_{sb}$ are excess conductivities at the points c and b due to the presence of nonequilibrium carrier concentration. In deriving the above equation the assumption was made $\Delta n=0$ outside the part b, c and $\Delta n=\Delta n_s=\text{const}$ inside the part b, c (Sec. 3.2 assumption ζ).

[72] J. Tauc, Phys. Rev. 95, 1394 (1954).

lated. For the determination of n_k one has 4 equations (1), (23), and (25) with 4 unknowns: $I_1=-I_2$, φ, n_1, n_2. It is necessary to introduce n_k into (1) instead of ζ_k using (8). One obtains

$$I_1=\sigma_1 F+eD_1\frac{dn_1}{dx}+\frac{\sigma_1}{e}Q_1^{**}\frac{1}{T}\frac{dT}{dx}, \tag{35}$$

$$I_2=\sigma_2 F-eD_2\frac{dn_2}{dx}-\frac{\sigma_2}{e}Q_2^{**}\frac{1}{T}\frac{dT}{dx}, \tag{36}$$

where

$$Q_k^{**}=Q_k^{*}-\langle\epsilon_k\rangle_{\text{Av}} \tag{37}$$

are the excess of the mean transfer kinetic energies Q_k^* over the mean kinetic energies

$$\langle\epsilon_k\rangle_{\text{Av}}=\frac{\int_0^\infty \omega_k(\epsilon_k)\epsilon_k e^{-\epsilon_k/\mathrm{k}T}d\epsilon_k}{\int_0^\infty \omega_k(\epsilon_k)e^{-\epsilon_k/\mathrm{k}T}d\epsilon_k}; \tag{38}$$

for spherical energy surfaces $\langle\epsilon_k\rangle_{\text{Av}}=\frac{3}{2}\mathrm{k}T$.

In some cases it will be sufficient to consider a schematical distribution of Δn for which $\Delta n=0$ outside the illuminated part b, c and $\Delta n=\Delta n_s=\text{const}$ inside this part. We call this assumption (ζ). It is approximately justified in case the illuminated part b, c is large compared with the diffusion length of excess current carriers and variation of the parameters of the semiconductor over a diffusion length is negligible; variation of τ and μ_k along the sample is neglected.

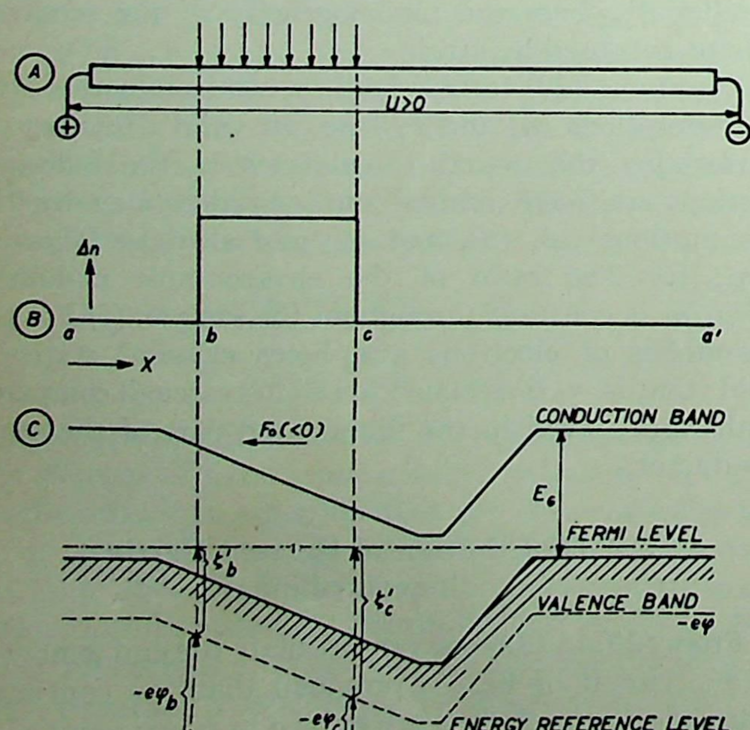


FIG. 5. Bulk photovoltaic effect: A the illuminated sample, B distribution of Δn, C energy levels in the dark.

4. BULK PHOTOVOLTAIC EFFECT

From the point of view of photovoltaic effects, the most important case of inhomogeneity of a semiconductor is an inhomogeneity in distribution of active impurity concentrations. We consider two limiting cases. In the first case the impurity concentration change occurs over a distance long compared with the diffusion length of current carriers; in the second case the change is abrupt. In the first case current carriers are in thermal equilibrium with the lattice; we use the word "bulk" to denote this type of photovoltaic effect. An example of the second case (barrier-layer photovoltaic effect) is the photovoltaic effect in a p-n junction (Sec. 5).

4.1 Basic Equations for the Bulk Photovoltaic Effect

Let us calculate the emf generated between the contacts a and a' in Fig. 1 for the case that in the semiconductor the active impurity concentrations are variable. Denoting the difference between the concentrations of donors and acceptors by $N=N_D-N_A$ we have $dN/dx\neq 0$ at least at some places between b and c. Such a bulk photocell is shown in Fig. 5.

The bulk photo emf is given by the integral (31); we shall perform the integration on assumptions (α), (β), (γ), (δ), and (ϵ). We substitute for $\Delta\zeta_k=\zeta_k-\zeta_{k0}$ from (8) and use the mass-action law

$$n_{10}n_{20}=n_i^2=\text{const}. \tag{39}$$

The emf $U=U_b$ can be expressed as a sum:

$$U=U_c+U_d, \tag{40}$$

where

$$U_c=\frac{\mathrm{k}T}{e}\oint\frac{\mu_1+\mu_2}{\mu_1 n_1+\mu_2 n_2}\frac{\Delta n}{n_{10}}\frac{dn_{10}}{dx}dx, \tag{41}$$

$$U_d=-\frac{\mathrm{k}T}{e}\oint\frac{\mu_1-\mu_2}{\mu_1 n_1+\mu_2 n_2}\frac{d\Delta n}{dx}dx. \tag{42}$$

If U_c is to be different from zero it is necessary that at least at some places between b and c $dn_{10}/n_{10}dx=1/\mathrm{k}T\cdot d\zeta_{10}/dx\neq 0$. This part U_c of the emf is caused by variation of the chemical potential of the assemblies of electrons and holes in the dark along the sample; we denote U_c as a "chemical" emf. The other part of the emf U_d corresponds to the diffusion voltage generated by the gradients of Δn in case electron and hole mobilities are different $\mu_1\neq\mu_2$. We denote U_d as a "diffusion" emf. These notations are analogous to the terminology used in the theory of galvanic cells.

An expression for U is obtained on the assumption of small illumination that is, if everywhere $\Delta\sigma\ll\sigma_0$. For an n-type semiconductor ($n_{10}=N$, $N\gg n_i$), we get

$$U=-\frac{\mathrm{k}T}{e}\frac{2}{1+b}\oint\Delta\sigma\frac{d\rho_0}{dx}dx. \tag{43}$$

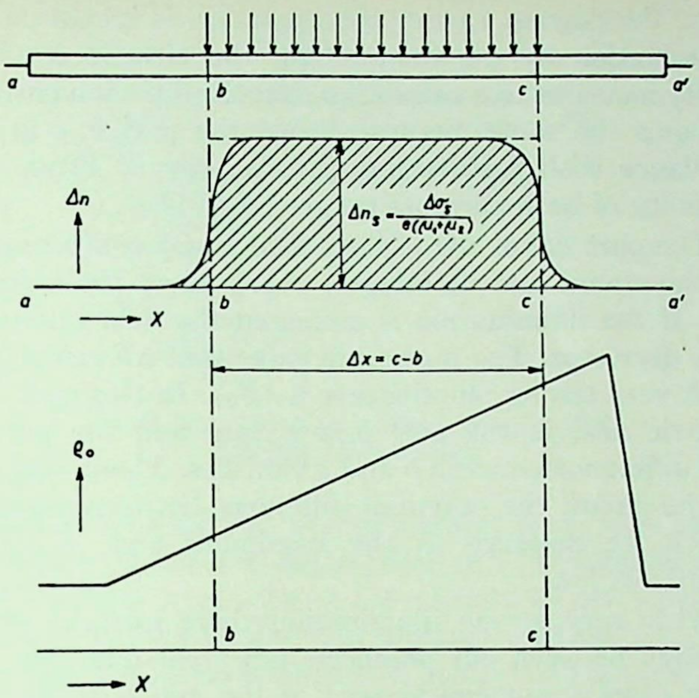


FIG. 6. Example of plot of $\Delta\sigma$ and ρ_0 for calculating the approximation for weak illumination.

An analogous expression for a p-type semiconductor ($n_{20}=-N, -N\gg n_i$) is obtained by changing the sign and putting $1/b$ instead of b. If in the illuminated part $d\rho_0/dx$ is constant then U is proportional to the shaded area as in Fig. 6.

To obtain an idea of the order of magnitude of a bulk emf under small illumination, let us consider a typical numerical example which can be easily realized with germanium: $b=2{,}1$; $\rho_{0c}-\rho_{0b}=15-5=10\ \Omega$ cm, $\sigma_{0,\text{av}}\approx 0,1\ \Omega^{-1}\,\text{cm}^{-1}$; $\Delta\sigma_s/\sigma_{0,\text{av}}=0,1$, $\Delta\sigma_s=10^{-2}\ \Omega^{-1}\,\text{cm}^{-1}$; $T=300°$K. From (43), we obtain $U=0.026\cdot 0.65\cdot 10\cdot 10^{-2}=1.7$ mv.

An equation for U which holds for arbitrary intensity of illumination (compatible of course with our assumptions) can be deduced if for the distribution of Δn assumption (ζ) (Sec. 3.2) is approximately valid. Equation (41) can be then written

$$U_c=\frac{kT}{e}\Delta\sigma_s\int_b^c\frac{d\sigma_{10}}{\sigma_{10}^2+\sigma_{10}\Delta\sigma_s+\sigma_{i1}\sigma_{i2}},\tag{44}$$

where $\sigma_{ik}=e\mu_k n_i$.

Writing $\mathfrak{D}=4\sigma_{i1}\sigma_{i2}-\Delta\sigma_s^2$, the result can be written:
(a) $\mathfrak{D}>0$, it is $\Delta\sigma_s<2(\sigma_{i1}\sigma_{i2})^{\frac{1}{2}}$ (very small illumination)

$$U_c=\frac{kT}{e}\Delta\sigma_s\frac{2}{\sqrt{\mathfrak{D}}}\times\text{arctg}\frac{2(\sigma_{10c}-\sigma_{10b})\sqrt{\mathfrak{D}}}{\mathfrak{D}+(2\sigma_{10c}+\Delta\sigma_s)(2\sigma_{10b}+\Delta\sigma_s)}.\tag{45a}$$

(b) $\mathfrak{D}<0$, it is $\Delta\sigma_s>2(\sigma_{i1}\sigma_{i2})^{\frac{1}{2}}$

$$U_c=-\frac{kT}{e}\Delta\sigma_s\frac{1}{\sqrt{-\mathfrak{D}}}\left(\ln\frac{2\sigma_{10c}+\Delta\sigma_s+\sqrt{-\mathfrak{D}}}{2\sigma_{10c}+\Delta\sigma_s-\sqrt{-\mathfrak{D}}}-\ln\frac{2\sigma_{10b}+\Delta\sigma_s+\sqrt{-\mathfrak{D}}}{2\sigma_{10b}+\Delta\sigma_s-\sqrt{-\mathfrak{D}}}\right).\tag{45b}$$

These expressions are simpler for an extrinsic semiconductor. For an n-type semiconductor we get from (45b) for arbitrary intensity of illumination

$$U_c=-\frac{kT}{e}\ln\frac{1+\Delta\sigma_s/\sigma_{0c}}{1+\Delta\sigma_s/\sigma_{0b}}.\tag{46}$$

Equation (42), giving the part U_d of the emf, can be written

$$U_d=-\frac{kT}{e}\frac{\mu_1-\mu_2}{\mu_1+\mu_2}\oint\frac{d\Delta\sigma}{\sigma_0+\Delta\sigma}.\tag{47}$$

Assuming the distribution (ζ) of Δn one has $d\Delta\sigma\neq 0$ only in the neighborhood of the points $x=b$, c; one easily obtains:

$$U_d=\frac{kT}{e}\frac{\mu_1-\mu_2}{\mu_1+\mu_2}\ln\frac{1+\Delta\sigma_s/\sigma_{0c}}{1+\Delta\sigma_s/\sigma_{0b}}.\tag{48}$$

This holds for a semiconductor of arbitrary type of conductivity.

For the total emf U in a n-type semiconductor, we obtain

$$U=-\frac{kT}{e}\frac{2}{1+b}\ln\frac{1+\Delta\sigma_s/\sigma_{0c}}{1+\Delta\sigma_s/\sigma_{0b}}.\tag{49}$$

The dependence of U on $\Delta\sigma_s$ is illustrated in Fig. 7 for several values of σ_{0c}/σ_{0b}. For very strong illumination $\Delta\sigma_s\gg\sigma_{0c}$ and $\Delta\sigma_s\gg\sigma_{0b}$ the emf approaches a saturation value given by

$$U_{\max}=-\frac{kT}{e}\frac{2}{1+b}\ln\frac{\sigma_{0b}}{\sigma_{0c}}.\tag{50}$$

In the above example $U_{\max}$ is equal to $0.026\cdot 0.65\cdot\ln 3=18.5$ mv.

4.2 Discussion

Let us consider the physical meaning of the "chemical" and "diffusion" parts of U. The generation of U_c

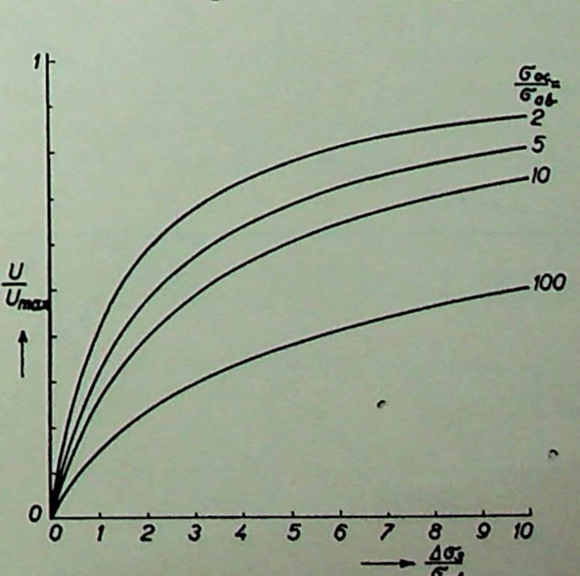


FIG. 7. Dependence of the relative photovoltage from (49) $U/U_{\max}$ on the intensity of illumination characterized by the ratio $\Delta\sigma/\sigma_{0b}$ for different σ_{0c}/σ_{0b}.

is conditioned by the difference in chemical potentials of electrons and holes in the dark at the boundary points of the illuminated part (in other words, by the difference of the Fermi levels relative to the boundaries of energy bands). We see this from (41) which for very strong illumination ($\Delta n_s \gg n_{10}, n_{20}$) on the validity of assumption (ζ) gives

$$U_{c\max}=\frac{1}{e}\int_b^c \frac{d\zeta_{10}}{dx}dx=(\zeta_c'-\zeta_b')/e. \qquad (51)$$

We can make the following detailed picture of the process. In Fig. 5, between the points b and c there exists in the dark a certain potential difference $(\zeta_b'-\zeta_c')/e$ because, in thermal equilibrium the electrochemical potential $\zeta'-e\varphi$ is everywhere constant. At every point between b and c there is an electric field F_0. The potential difference between the points b and c is compensated by a potential difference of opposite sign so between the points a and a' the potential difference is zero in the dark. If the condition that the two ends of the semiconductor have the same properties is not fulfilled at least a part of the compensating voltage is situated at the contacts.

If in the part b, c excess current carriers are generated by illumination, the electric field F_0 sets them in motion. When the contacts a and a' are not connected an electric charge is created which causes such a change in the intensity of the field F_0 that the current is annulled. In our case, (Fig. 5) the contact a is charged positively, the contact a' negatively (with our sign convention $U>0$).

The field intensity in the part b, c decreases by illumination from the dark value $|F_0|$ to the value $|F|=|F_0-\Delta F$ with $\Delta F>0$. At every point in part b, c the electric current density produced by the electric field $\sigma_0 F_0$ compensates that produced by diffusion. If, now, the current carrier concentration is increased by illumination by an amount Δn, the electric conductivity increases to a value $\sigma>\sigma_0$ but the diffusion current remains the same, because inside the part b, c in accordance with assumption (ζ) $d\Delta n/dx=0$. From the equality of both currents we get $|F|<|F_0|$.

The part ΔF of the electric field is not compensated in the circuit and its integral over x gives the voltage U_c. If the illumination is increased the field intensity $|F|$ decreases. The maximum value that ΔF can attain with very strong illumination is $|F_0|$. In this case the electric field in the part b, c is zero and the potential difference between b and c vanishes. There remains in the circuit the potential difference between c and b which we measure as the maximum emf, $U_{c\max}=-(\zeta_b'-\zeta_c')/e$.

With very strong illumination there exists a close analogy between our photocell and a galvanic cell; in both cases the chemical part of the emf corresponds to the transfer of an electron from the point with a chemical potential ζ_c' to a point with a chemical potential ζ_b'. With weak illumination there is a difference between a photocell and a galvanic cell as the emf of a photocell is a function of the illumination. This is due to the fact that in the photocell the emf is generated only by nonequilibrium current carriers while the conductance of the photocell is determined by all current carriers. When the concentration of excess carriers is very much larger than the concentration of dark carriers the situation is similar to that in a galvanic cell where all current carriers contribute to the generation of the emf. One has a similar case in the photocell proposed by Kramer[73] and later improved by Ohmart.[74] This photocell is constructed of two metal plates of different work functions with a gas between them. The gas is ionized by absorption of radiation. It was shown by experiment that the emf is determined by the contact potential of the metals and is independent of illumination intensity which changes only the internal resistance. Other examples are the photocells described by Nadjakov and Andreitchin[75] which consist of two different metal plates separated by sulfur or other insulator. As the internal resistance is very high, the emf attains its saturation value at low illumination. These examples correspond well with the general point of view on the generation of a "chemical" emf explained in this section.

The other part is the diffusion emf U_d. An electric field arises in those parts of the semiconductor where $d\Delta n/dx\neq 0$ for the case that the electron and hole mobilities are not equal. The direction and intensity of this field are such as to accelerate the motion of current carriers with lower mobility and to slow down the motion of current carriers with larger mobility in order

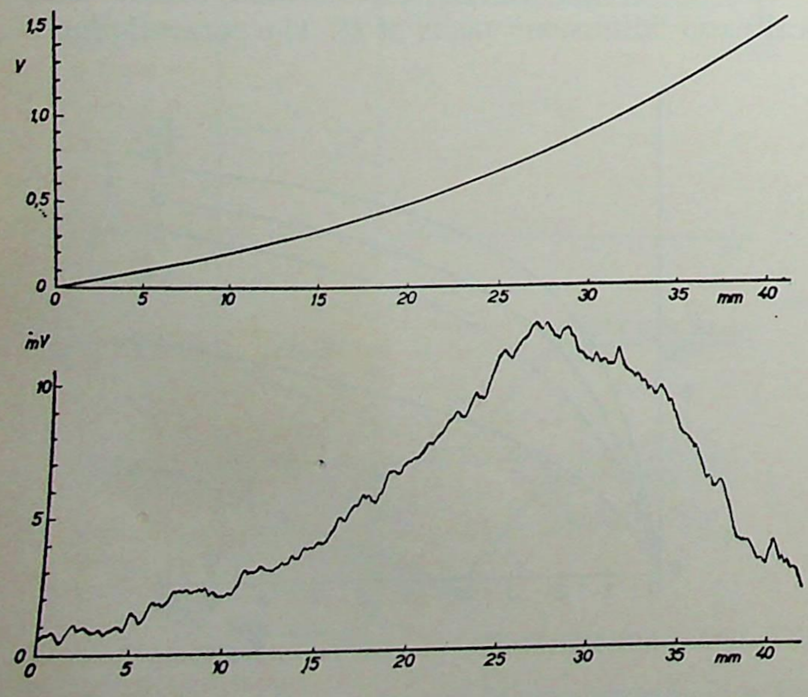


FIG. 8. Example of Trousil's measurement of the bulk photovoltage.

[73] J. B. Kramer, Electrician 93, 497 (1924).
[74] P. E. Ohmart, J. Appl. Phys. 22, 1504 (1951).
[75] G. Nadjakov and R. Andreitchin, Compt. rend. acad. bulgare sci. 5, 9 (1952); *ibid.*, 7, 13 (1954); Izvest. Ban, Ser. Fiz. 2, 293 (1951).

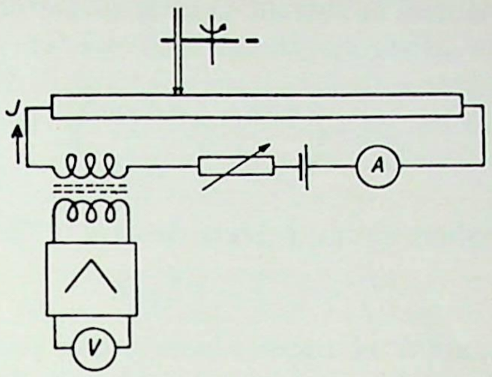


FIG. 9. Compensation method for measuring the bulk photovoltage.

to make the resultant electric current zero. It is usual to denote this kind of electric field as a "Dember field." The voltages generated in the vicinity of both points b, and c [where $d\Delta n/dx \neq 0$ according to assumption (ζ)] are mutually compensated in a homogeneous semiconductor. This is not the case if at the points b, and c the conductivities are not the same [see Eq. (48)] and the resultant emf U_d is measured between a, and a'. This emf is analogous to the diffusion emf generated in a galvanic element between two electrolytes. Just as with galvanic cells it is possible to design semiconductor photocells in which $|U_d|$ is small compared with $|U_c|$.

4.3 Comparison with Experiment and Applications

The first experimental investigation[45,76] of the theory of the bulk photovoltaic effect was carried out by Trousil[46] on long samples cut from germanium single crystals drawn from the melt. An example of Trousil's measurement is given in Fig. 8. In the upper figure there is shown the voltage V measured along the sample in the dark with a current flowing through it; the resistivity $\rho_0 \sim dV/dx$. The lower figure shows the emf generated by a spot of light moved along the sample and measured with a recording galvanometer. Qualitative agreement with the theory is apparent, e.g., the maximum of the emf is situated at the point of the highest value of d^2V/dx^2. From the emf curve small inhomogeneities in the distribution of the concentrations of impurities can be detected even if they are too small to be observable by measurement of conductivity along the sample.

Frank [77] used this effect to work out a highly sensitive compensation method for the determination of inhomogeneities of semiconductor single crystals. His experimental arrangement is shown schematically in Fig. 9; here the bulk emf is compensated by the photoconductive signal $J\delta R$, where δR is the change of the circuit resistance by illumination. If the current density J/Q (Q is the cross section of the sample) is so chosen that the voltmeter has zero deviation then for an n-type semiconductor

$$\frac{d\sigma_0}{dx}=\frac{e}{kT}\,\frac{2}{b+1}\,\frac{J}{Q}. \tag{52}$$

Using this method Frank made a quantitative comparison of the theory[45] of the bulk photovoltaic effect and obtained good agreement. Recently Lashkarev and Romanov[78] have published a paper on the bulk photovoltaic effect investigated with a very narrow light spot.

5. BARRIER-LAYER PHOTOVOLTAIC EFFECT

5.1 Basic Equation of a *p-n* Junction Photovoltaic Cell

Between a p-type and an n-type semiconductor there is, in the dark, a potential barrier of height $\varphi_{b0}=|\zeta_P'-\zeta_N'|$. This barrier can be made very thin and the electric field within it can attain such high values that the change in potential energy over a distance of the same order of magnitude as the mean free path is comparable with the mean thermal energy of an electron or greater. Then (1) is no longer valid; inside such a barrier some quantities such as chemical potentials lose their meaning. However, it is easy to describe another limiting case when the thermal energy of an electron is negligible compared with the energy the electron acquires over the mean free path. It is then usual to speak of the so-called "diode" theory to stress the fact that, in analogy with the situation met with in vacuum tubes, the electron motion is determined primarily by electric fields and influenced little by electron-ion or electron-electron collisions. Of course, this theory holds only in a limited space, in a p-n junction or in its immediate neighborhood. In other regions (1) is valid. It is thus possible to use (1) and respect the regions where the electron motion is described by the diode theory by convenient boundary conditions.

Let us consider the photocell shown in Fig. 10. The contacts here are again connected to the semiconductor of the same properties. The left p-n junction is illuminated. The illumination extends on both sides of the junction over distances which are large compared with carrier diffusion lengths. The p- and n-type semiconductors are assumed to be homogeneous (n_{k0}=const). The potential distribution is drawn in the lower figure.

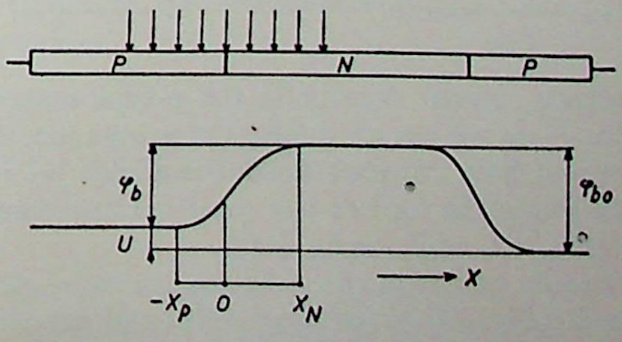


FIG. 10. p-n junction photocell. For the calculation the cross section of the semiconductor is assumed to be unity.

[76] J. Tauc, Czechosl. J. Phys. 5, 300 (1955).
[77] H. Frank, Czechosl. J. Phys. 6, 433 (1956).
[78] V. E. Lashkarev and V. A. Romanov, Trudy Inst. fiz. Akad. Nauk USSR, Kiev 7, 50 (1956).

The diode theory holds between $-x_P$, x_N. Calculation of φ_{b0} and x_P, x_N can be based[57,58] on (26). The barrier thicknesses x_P+x_N are assumed to be small compared with the diffusion lengths of carriers so recombination inside the barrier can be neglected. In regions outside the barrier we assume that $n_{2N} \ll n_{1N}$, $n_{1P} \ll n_{2P}$ (not very great illumination). Then the electric fields are small and the current densities are determined by diffusion. If assumptions (α), (β), (γ), (δ), and (ϵ) are valid we get from (35) and (36) for current densities

$$I_k = \pm kT\mu_k \frac{dn_k}{dx} \quad \begin{matrix} + \text{ for } k=1, \\ - \text{ for } k=2. \end{matrix} \tag{53}$$

We consider holes in the n-type material. Substituting from (53) into (23) we get the following equation for $\Delta n = n_k - n_{k0}$:

$$\frac{d^2\Delta n}{dx^2} - \frac{\Delta n}{L_{2N}^2} = -\frac{g_f\tau}{L_{2N}^2}. \tag{54}$$

For boundary conditions we take

$$x=\infty, \quad \Delta n = 0, \tag{55}$$

$$x = x_N, \quad n_{2N} = n_{2P}e^{-e\varphi_b/kT}, \tag{56}$$

where φ_b is the barrier height. The boundary condition (56) expresses the ratio of the concentrations of holes on the two sides of the barrier by the Boltzmann factor as required by classical statistics. We have assumed further that the hole concentration at the point $-x_P$ is so great that its diminution by the diffusion of holes into the n-type material is negligible; we thus have at the point $-x_P$, $n_{2P} = n_{2P0}$.

It is easy to find the solution Δn of (54) with boundary conditions (55) and (56). Putting this solution into (53) we get for the current density

$$I_2 = \frac{kT\mu_2}{L_{2N}}(n_{2P0}e^{-e\varphi_b/kT} - n_{2N0} - \tau g_f)e^{-x-x_N/L_{2N}}. \tag{57}$$

In the dark, for the barrier height φ_{b0}, from (56)

$$n_{2N0} = n_{2P0}e^{-e\varphi_{b0}/kT}. \tag{58}$$

Using φ_{b0} we obtain for the current density at the point x_N:

$$I_{2N} = \frac{kT\mu_2}{L_{2N}} n_{2N0}(e^{-e(\varphi_b-\varphi_{b0})/kT} - 1) - \frac{kT\mu_2}{L_{2N}} g_f\tau. \tag{59}$$

For electron current density in the p-type material at the point $-x_P$ we get an analogous expression. As the electron and hole currents do not diminish by recombination inside the barrier the total current density is $I = I_{1N} + I_{2N} = I_{1P} + I_{2P} = I_{2N} + I_{1P}$.

The barrier height in the dark φ_{b0} is compensated in the circuit (Fig. 10) so for measurable voltage we have $U = -(\varphi_b - \varphi_{b0})$. Thus, we obtain the basic equation for the photocell considered

$$I + I_f = I_s(e^{eU/kT} - 1). \tag{60}$$

I_s is the saturation current density of the photocell in the dark, i.e., I for $I_f = 0$ and $-U \gg kT/e$:

$$I_s = kT\left(\frac{\mu_1 n_{1P0}}{L_{1P}} + \frac{\mu_2 n_{2N0}}{L_{2N}}\right). \tag{61}$$

$-I_f$ is the short circuit current density of the photocell

$$I_f = eg_f(L_{1P} + L_{2N}). \tag{62}$$

I_f is the number of excess electron-hole pairs per cm² reaching the region of the potential barrier by diffusion per sec.

The derivation of formula (60) is analogous to that used by Shockley[57] in his theory of p-n junction rectifiers. Expression (60) reduces to Shockley's equation (4.22) for $I_f = 0$. For other constructions of p-n junction photocells several formulas are given in the literature.[79,80]

The emf which follows from (60) with $I = 0$ is

$$U_c = \frac{kT}{e}\ln\left(1 + \frac{I_f}{I_s}\right). \tag{63}$$

An expression of this form was deduced by Fan[81] and experimentally verified by Becker and Fan.[23] It follows from assumptions on which (63) was derived that U_c cannot exceed the height of the barrier in the dark φ_{b0} and must saturate at very high illumination when (63) is no longer valid. When U_c approaches φ_{b0}, the diode theory ceases to hold as the electric field in the barrier is diminished; we then have a transition to the bulk photovoltaic effect—a problem which has not yet been studied.

We have denoted the emf expressed by (63) as U_c because it actually corresponds to the difference in the chemical potentials and may therefore be denoted as a "chemical" emf. Calculation of the diffusion emf U_d is the same as given in Sec. 4. Under the conditions assumed in the present section and at small illumination we get approximately

$$U_d = \frac{kT}{e}\frac{b-1}{b+1}(\rho_{0P} - \rho_{0N})\Delta\sigma_s. \tag{64}$$

The p-n junction is usually constructed so that p- and n-type materials have high conductivities so that $\Delta\sigma_s/\sigma_{0P} \ll 1$, $\Delta\sigma_s/\sigma_{0N} \ll 1$ and $|U_d|$ is usually negligible compared with U_c.

5.2 Comparison of the Barrier-Layer and the Bulk Photovoltaic Effects

The bulk and the barrier-layer photovoltaic effects are two limiting cases of one and the same physical phenomenon; in both cases the emf is generated by an

[79] J. Tauc, Československý časopis pro fysiku — printed: J. Tauc, Českosl. čas. fys. 4, 158 (1954).
[80] R. L. Cummerow, Phys. Rev. 95, 16 (1954).
[81] H. Y. Fan, Phys. Rev. 75, 1631 (1949).

internal electric field and only the details of the process are different. The first case approaches the thermal equilibrium state, the other is radically different from it. If we denote the time an electron spends in the active space of the photocell τ_u, then for a barrier layer photocell $\tau_u \ll \tau$, for a bulk photocell $\tau_u \gg \tau$.

The emf U_c with very strong illumination attains in both cases the same saturation value $(\zeta_c'-\zeta_b')/e$. The maximum value this expression can have in a given semiconductor is approximately the value E_G/e. This is true for the case that at the point b the material is highly doped with p-type impurities and at the point c with n-type impurities. Even if the maximum emf is the same for both effects the short circuit current of a barrier layer photocell is usually much higher for the appropriate geometry of illumination than that of a bulk photocell. Consequently, the barrier layer photocell transforms radiation energy into electrical energy with a higher efficiency than a bulk photocell. Indeed, in the former case all electrons and holes created in the active space of the photocell take part in the generation of the short circuit current while in the latter case the short circuit current is diminished by recombination. We shall make a rough quantitative comparison of both cases.

Let us consider a p-n junction photocell of unit cross section which is transversely illuminated with a narrow light spot so that all incident photons are absorbed in the potential barrier space. With quantum efficiency equal to unity these N_f photons per second give rise to a short circuit current $|I_{sc}^{\text{barrier}}|=eN_f$. Let the same number N_f photons/sec be absorbed in a bulk photocell of unit cross section; the illumination takes place over a distance $c-b=\Delta x$. The added carrier concentration is $\Delta n_s = g_f\tau = N_f\tau/\Delta x$ and the short circuit current $|I_{sc}^{\text{bulk}}|=|U|/R$, where U is the bulk photo emf (Sec. 4.1) and R the resistance of the illuminated photocell. For strong illumination ($\Delta n_s \gg n_{10}$ and n_{20}), we obtain for the ratio of both short circuit currents

$$|I_{sc}^{\text{bulk}}|/|I_{sc}^{\text{barrier}}|=\tau/\tau', \tag{65}$$

where $\tau'=\Delta x/(\mu_1+\mu_2)F'$, $F'=U/\Delta x$. Putting typical values for germanium into this formula ($\mu_1+\mu_2=1700+3600=5300$ cm²/vs, $U_{\max}\approx 0.6$ v, $\tau=10^{-4}$ sec) and choosing $\Delta x=2$ cm ($\gg L_1, L_2$) we get $\tau'=10^{-3}$ sec and $|I_{sc}^{\text{bulk}}|=0,1|I_{sc}^{\text{barrier}}|$.

However, the bulk photocell can be more advantageous from the point of view of efficiency than a barrier layer photocell for the case that the radiation cannot be concentrated in the barrier layer. This can be the case for high energy radiation which is little absorbed in the semiconductor and cannot be concentrated by optical means. Let us consider a case that the barrier layer photocell is irradiated over distances large compared with L_{1P}, L_{2N} on both sides of the barrier. This case was treated in Sec. 5.1; from (62) we obtain $|I_{sc}^{\text{barrier}}|=I_f=e(L_{1P}+L_{2N})g_f$. Then we have

$$\frac{|I_{sc}^{\text{bulk}}|}{|I_{sc}^{\text{barrier}}|}=\frac{\tau}{\tau'}\frac{\Delta x}{L_{1P}+L_{2N}}. \tag{66}$$

In the numerical example considered, $L_{1P}=9.6\cdot 10^{-2}$ cm, $L_{2N}=6.6\cdot 10^{-2}$ cm, $|I_{sc}^{\text{bulk}}|=1.25|I_{sc}^{\text{barrier}}|$.

We shall add some general remarks concerning the energy conversion efficiency by a photocell. From the point of view of thermodynamics there is no reason for this efficiency not to approach 100%. Of course this is possible only for a photocell so constructed that $|U_{\max}|=E_G/e$ and illuminated by monochromatic radiation of wavelength hc/E_G. With a radiation of shorter wavelength the efficiency is always smaller; with a high energy radiation (Sec. 1.3) the maximum efficiency can be only E_G/ϵ. An ideal photocell should have an infinite dark resistance [see (63)]; it is desirable to have I_s very small so that $|I|\approx I_f$ even if $U\neq 0$ and zero resistance at illumination for the internal losses to be small.

5.3 Classification of Simple Photovoltaic Effects

As a consequence of the preceding considerations, we can divide photovoltaic effects into barrier layer and bulk effects. We can further distinguish between effects at the contacts or inside a semiconductor crystal.

It is necessary to consider how the well-known Dember effect can be put into this classification. The Dember effect is the generation of an emf in a homogeneous semiconductor without any barrier layers. We consider here, as everywhere in the present paper, only the stationary case and exclude therefore the Dember effect observed with interrupted illumination.[82–86] The inhomogeneity of the circuit necessary from our point of view for the generation of a stationary emf is realized here by the fact that a part of the circuit d, d' is of semiconductor, the other part $(a,d;\, d',a')$ of metal as shown in Fig. 11. The nonrectifying semiconductor-metal contact is essential for the stationary Dember effect. Table I contains the proposed classification of simple photovoltaic effects.

In the present paper we do not consider generation of an emf at semiconductor-metal contacts. This is more difficult to discuss in detail owing to the more complicated structure of these contacts that is not yet quite understood.

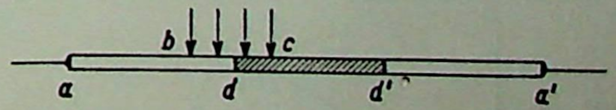


FIG. 11. Dember effect.

[82] L. Bergmann, Physik. Z. 33, 209 (1932).
[83] L. Bergmann, Z. Physik 100, 50 (1936).
[84] S. M. Ryvkin, Zhur. Tech. Fiz. 18, 1521 (1948); 19, 286 (1949).
[85] K. B. Tolpyvo, Izvest. Akad. Nauk, Ser. Fiz. 16, 46 (1952).
[86] Moss, Pincherle, and Woodward, Proc. Phys. Soc. (London) B66, 743 (1953).

TABLE I. Classification of simple photovoltaic effects.

Photovoltaic effects	Barrier layer	Nonbarrier layer
Noncontact	In p-n junction photocells	Bulk photovoltaic effect
Contact	In barrier-layer photocells (such as cuprous oxide photocells) In point contact photocells	One contact illuminated: Dember effect Both contacts illuminated: Effects observed by Nadjakov and Andreitchin

6. SOME MORE COMPLICATED PHOTOVOLTAIC EFFECTS

An essential condition for the generation of an emf in a semiconductor, following the preceding considerations, is a chemical potential difference between the boundaries of the illuminated part b, c. In cases dealt with in Sec. 4 and 5 this difference was realized by a donor or acceptor concentration gradient. In this section we discuss two other examples of this rule when the chemical potential difference is realized by a magnetic field or a temperature gradient.

6.1 Photoelectromagnetic Effect in a Nonhomogeneous Magnetic Field[87]

Landau[88] was the first to show that the chemical potential of an assembly of free electrons depends on the intensity of the magnetic field. Therefore, if the magnetic field H situated in a plane perpendicular to the axis of the sample at the point b differs from that at the point c an emf is generated between the contacts a, and a'. Using Landau's expression for the thermodynamic potential of the assembly of free electrons we obtain from (31) for the case shown in Fig. 12 the following equation for U_c (with validity of assumptions (α), (β), (γ), (δ), (ϵ), and (ζ) and for $|H| \ll kTm_k/\beta m$):

$$U_c = -\frac{kT}{e}\Delta t_1\left\{1-\frac{1}{6}\left[\left(\frac{m}{m_1}\right)^2+\left(\frac{m}{m_2}\right)^2\right]\right\} \times \frac{\beta^2(H_c^2-H_b^2)}{k^2T^2}. \quad (67)$$

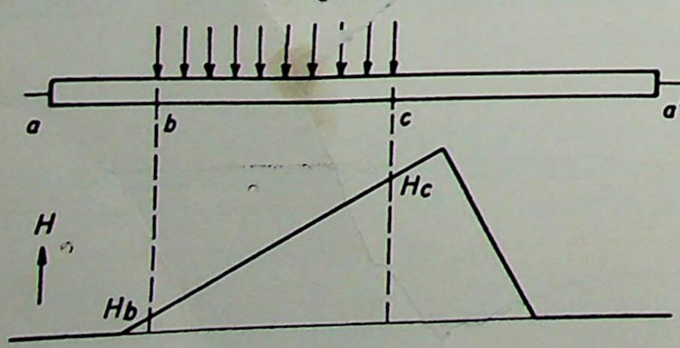


FIG. 12. Photoelectromagnetic effect in a nonhomogeneous magnetic field.

[87] J. Tauc, Czechosl. J. Phys. 6, 421 (1956).
[88] L. Landau, Z. Physik 64, 629 (1930).

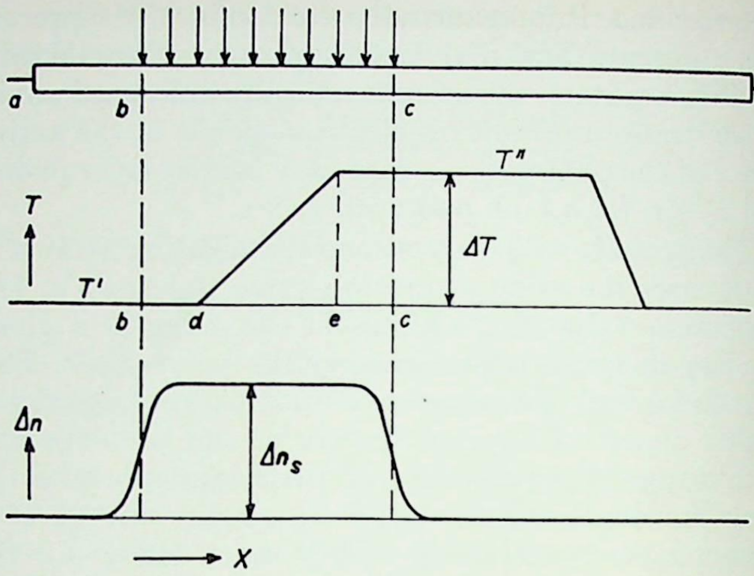


FIG. 13. Scheme of the photothermoelectric effect.

We have assumed spherical energy surfaces in the Brillouin zone and used the effective mass concept for electrons and holes. β denotes the Bohr magneton, $\Delta t_1 = t_1 - t_{10}$.

This emf differs from that observed by Kikoin and Noskov[89] and others, [90,91] in a homogeneous magnetic field. In the case of Kikoin and Noskov's photoelectromagnetic effect the emf is generated perpendicularly to the magnetic field and to the direction of illumination. It differs also with regard to dependence on magnetic field, which for small fields is linear for the Kikoin-Noskov effect and quadratic for U_c as given by (67). If the direction of the magnetic field is parallel to the direction of illumination the Kikoin-Noskov emf should be theoretically zero but the emf U_c given by (67) is not.

The experimental determination of this emf should be interesting as it could give us some information on effective masses just as cyclotron resonance or magnetic susceptibility measurement. In some materials, such as InSb, U_c is expected to attain measurable values. But this emf is masked by a much stronger emf generated at the boundaries of the illuminated part in a nonhomogeneous magnetic field which acts on the diffusion currents of electrons and holes. With assumptions (α) to (ζ), for small illumination $\Delta\sigma_s/\sigma_0 \ll 1$, small induction B of the magnetic field ($B\mu_2 \ll 1$) and an n-type semiconductor we obtain for this emf the approximate expression

$$U_e = \frac{kT}{e}\frac{\Delta\sigma_s}{\sigma_0}\mu_2^2(B_c^2-B_b^2). \quad (68)$$

This emf was actually found experimentally[87] in InSb.

[89] J. K. Kikoin and M. M. Noskov, Physik. Z. Sowjetunion 5, 586 (1934).
[90] P. Aigrain and H. Bulliard, Compt. rend. 236, 595 and 672 (1953).
[91] H. Bulliard, Ann. phys. 15, 52 (1954).

6.2 Photothermoelectric Effect

Another way[92] of realizing the chemical potential difference is by putting the points b, c at different temperatures T', T''. The principle of measurement of this effect is given in Fig. 13. The normal thermoelectric voltage is not generated in the circuit as the points a, and a' are at the same temperature. An experimental arrangement is illustrated in Fig. 14. A thin semiconductor sample is in good thermal contact with two isolated metal blocks and is illuminated with interrupted light. If the temperature of the blocks is different a signal can be measured by means of an alternating current amplifier and a valve voltmeter; the normal thermoelectric voltage is thus eliminated in this experimental arrangement.

The expression for the emf originally given[92] was generalized by Van der Pauw and Polder[93] who took into account the temperature dependence of E_G and b. Their expression is readily deduced from (29) and (30)

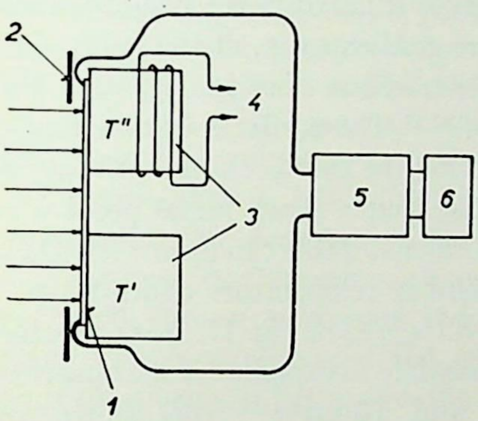


FIG. 14. Experimental arrangement for observing the photothermoelectric effect. 1. The illuminated sample. 2. Diaphragms screening the contacts. 3. Anodized aluminium blocks. 4. Heating of the upper block. 5 and 6. Narrow band amplifier with valve voltmeter.

with the usual assumptions (α), (β), (γ), (ϵ), and (ζ) and $\Delta T = T'' - T' \ll T'$, $\Delta\sigma_s/\sigma_b \ll 1$:

$$U_{PTE} = \frac{k\Delta T}{e}\left\{\frac{b-1}{b+1} + \frac{b}{b+1}\frac{N}{bn_{10}+n_{20}} \times\left[\frac{Q_1^{**}+Q_2^{**}}{kT} + \frac{1}{k}\frac{dE_G}{dT}\right] - \frac{1}{b+1}\frac{T(n_{10}+n_{20})}{bn_{10}+n_{20}}\frac{db}{dT}\right\}\frac{\Delta\sigma_s}{\sigma_0}. \quad (69)$$

From the measurement of this emf the sum $Q_1^{**}+Q_2^{**} = Q_1^{*}+Q_2^{*} - \langle\epsilon_1\rangle_{Av} - \langle\epsilon_2\rangle_{Av}$ can be determined. Under favorable circumstances the determination of this sum may give certain information on the carrier scattering mechanism (definitions of Q_k^* and $\langle\epsilon_2\rangle_{Av}$ in Secs. 2.3 and 3.2).

7. OTHER CASES OF GENERATION OF AN emf

Photoelectric generation of nonequilibrium current carrier concentration is not essential for the production of an emf in semiconductors. A nonequilibrium carrier

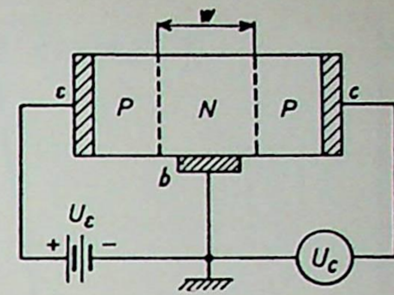


FIG. 15. Shockley's p-n-p junction.[57]

concentration can be generated by other means as well. We shall now deal briefly with some examples. Their theory can be developed from the same basic ideas as in the preceding cases, only the geometry and consequently carrier concentration distribution is usually more complicated.

A very important method for the production of excess carrier concentration is current injection which forms the basis for the functioning of the transistor. Bardeen considers[94] how holes injected by a point contact emitter into n-type germanium produce an emf at another point contact (collector) situated in the neighborhood of the first. His theory is shown to be in agreement with experiments performed by Pearson. Shockley[57] deals with a basically similar case, with the difference that the emitter and the collector are formed by a p-n junction. The arrangement Shockley considers is illustrated in Fig. 15. For U_c he deduces the following expression (on the assumption that $U_\epsilon > kT/e$):

$$U_c = U_\epsilon - \frac{kT}{e}\ln\cosh\frac{2w}{L_{2N}}. \quad (70)$$

L_{2N} is the hole diffusion length in n-type material, U_ϵ is the emitter voltage.

An emf may also be generated[95,96] in the case that nonequilibrium carrier concentration is produced in sharp temperature gradients. For a temperature drop concentrated within a distance shorter than the carrier diffusion length a nonequilibrium concentration is produced which is positive ($\Delta n > 0$) in regions of lower temperature and negative ($\Delta n < 0$) in regions of higher temperature. For instance, for an n-type semiconductor for the case that the temperature drop is concentrated

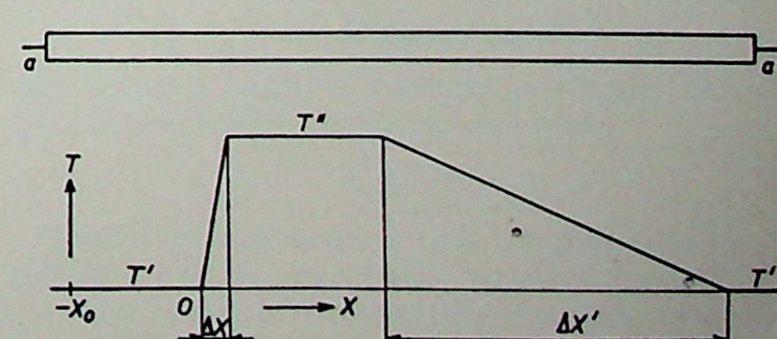


FIG. 16. Principle of measuring the Benedicks effect.

[92] J. Tauc, Czechosl. J. Phys. 5, 528 (1955).
[93] L. J. van der Pauw and D. Polder, J. Electronics 2, 239 (1956).
[94] J. Bardeen, Bell System Tech. J. 29, 469 (1950).
[95] J. Tauc, Czechosl. J. Phys. 3, 282 (1953).
[96] J. Tauc, Czechosl. J. Phys. 6, 108 (1956).

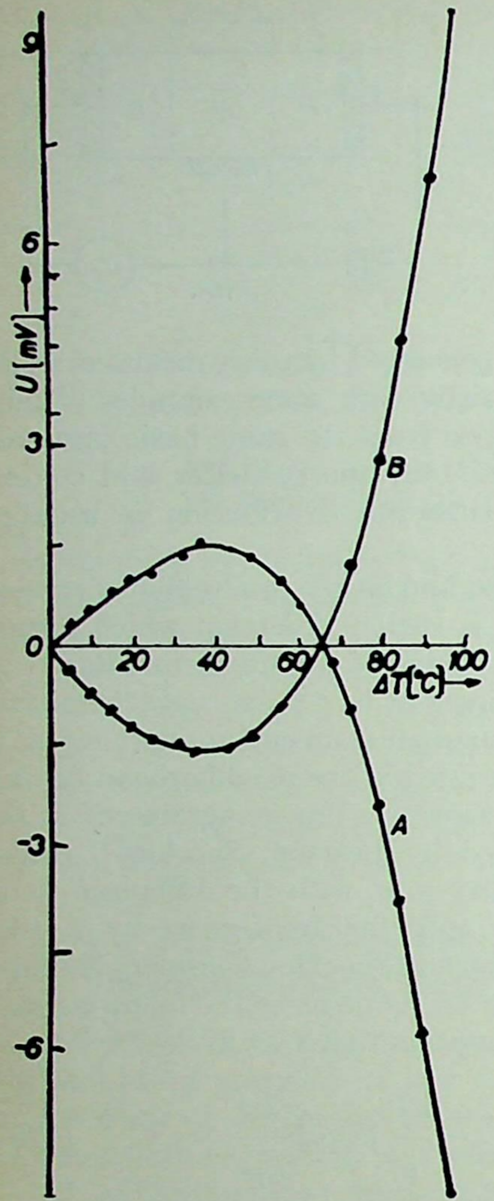


FIG. 17. Trousil's measurement of the Benedicks effect in germanium. The curves A and B correspond to both directions of the temperature drop.

within a distance shorter than the hole diffusion length one obtains (assuming $|\Delta n| \ll bn_{10}$):

$$\Delta U = \frac{k\Delta T}{e}\left[\frac{Q_1^* + Q_2^*}{kT} + (b-1)\right]\frac{n_{20}(T'') - n_{20}(T')}{2bn_{10}}. \quad (71)$$

We have denoted this emf as ΔU to emphasize that it adds to the normal thermoelectric voltage and causes its dependence on the temperature gradient—the so-called Benedicks effect. In the circuit shown schematically in Fig. 16 both the ends a, a' are at the same temperature and the normal thermoelectric voltage is zero. In the semiconductor two temperature drops of opposing sign are present, the right one occurring over a distance much longer than the diffusion length (so that here $\Delta U = 0$), the left one at a distance shorter than the diffusion length. Under these conditions the emf measured between a, a' is ΔU. This experiment[95] was realized by Trousil.[97] The result is shown in Fig. 17. As the sample was not quite homogeneous a small ordinary thermoelectric voltage was also generated. However, owing to strong temperature dependence of n_{20} in (71), ΔU rapidly increases with temperature difference and is much greater than the ordinary thermoelectric voltage at higher temperatures T''. This effect is much more easily measurable than the Benedicks effect in metals[98,99] as it is much greater.

If a potential barrier is situated in the neighborhood of a sharp temperature gradient, e.g., at the point $-x_0$ in Fig. 16 an emf is generated in close analogy with the experiments described.[94,57] These effects were actually observed when measuring the thermoelectric voltage in germanium for the case that a sharp metal point was used as one contact. Under suitable conditions one then gets a qualitatively similar temperature dependence[100] of the thermoelectric voltage as in Fig. 17. It is probable that in this way it is possible to explain[101] some observations by Granville and Hogarth[102] concerning an anomalous temperature dependence of the thermoelectric voltage measured with a metal point contact on polished surfaces of Ge or PbS crystals.

[97] Z. Trousil, Czechosl. J. Phys. **6**, 170 (1956).
[98] C. Benedicks, Erg. exakt. Naturw. **8**, 26 (1929).
[99] G. Kocher, Ann. phys. **16**, 210 (1955).
[100] J. Tauc and Z. Trousil, Czechosl. J. Phys. **3**, 120 (1953).
[101] J. Tauc, Czechosl. J. Phys. **3**, 259 (1953).
[102] J. W. Granville and A. C. Hogarth, Proc. Phys. Soc. (London) **B64**, 488 (1951).

Three Lectures on Relativity Theory

V. Fock

Physical Institute, University of Leningrad, U.S.S.R.

(Delivered to the Colloquium of the Institute for Theoretical Physics, Copenhagen, Denmark, February 18, 20, 22, 1957.)

FIRST LECTURE: ON HOMOGENEITY, COVARIANCE, AND RELATIVITY

In my first lecture, I try to elucidate some general notions connected with relativity theory. I speak on homogeneity, covariance, and relativity. My considerations are of a very simple nature but, nevertheless, I hope that they may be of interest, because simple notions are often the most difficult ones.

If we consider the geometrical aspect of the theory of space and time, this theory naturally divides into the theory of homogeneous (uniform) space-time and that of the nonhomogeneous (nonuniform) space-time. The former may be called Galilean space and the latter the Riemannian or Einsteinian space. (I sometimes use the word space instead of space-time.)

The property of space-time of being homogeneous means that (a) there are no privileged points in space and in time; (b) there are no privileged directions, and (c) there are no privileged inertial frames (that all frames moving uniformly and in a straight line with respect to one another are on the same footing).

The uniformity of space and time manifests itself in the existence of the Lorentz group. In particular, the equality of points in space and time corresponds to the possibility of a displacement, the equality of directions corresponds to that of spatial rotations, and the equality of inertial frames corresponds to a special Lorentz transformation. The displacements contain four parameters, the rotation three (the three angles), and the transformation to a moving frame also three (the three components of velocity). This gives together ten parameters—the maximum possible number, if we do not take into account scale transformations $x'=\lambda x$.

The statement that the Lorentz transformation leaves invariant the expression for the square of the line element is to be understood in the following sense.

If one writes ds^2 as

$$ds^2=dx_0^2-dx_1^2-dx_2^2-dx_3^2$$

or

$$ds^2=\eta_{\mu\nu}dx_\mu dx_\nu$$

then, after the transformation from (x) to (x'), we have

$$ds^2=\eta_{\mu\nu}dx_\mu' dx_\nu'$$

with *the same* matrix

$$\|\eta_{\mu\nu}\|=\begin{Bmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{Bmatrix}.$$

In studying the properties of homogeneous (uniform) space-time, the use of Galilean coordinates is convenient, but not essential. The property of space-time of being uniform may be as well expressed in general coordinates.

Let the substitution

$$x_\mu'=f_\mu(x_0x_1x_2x_3)\equiv f_\mu(x)$$

be performed in the expression for ds^2. Then,

$$ds^2=g_{\mu\nu}(x)dx_\mu dx_\nu$$

changes into

$$ds^2=g_{\mu\nu}'(x')dx_\mu' dx_\nu'$$

so that

$$g_{\mu\nu}'(x')=g_{\alpha\beta}\frac{\partial x_\alpha}{\partial x_\mu'}\frac{\partial x_\beta}{\partial x_\nu'}. \quad (1)$$

If the mathematical form of the functions $g_{\mu\nu}'$ is the same as that of the $g_{\mu\nu}$, that is, if

$$g_{\mu\nu}'(x')=g_{\mu\nu}(x), \quad (2)$$

then the space admits a transformation group.

For an infinitesimal transformation

$$x_\alpha'=x_\alpha+\eta^\alpha(x)$$

this leads to

$$\nabla^\mu\eta^\nu+\nabla^\nu\eta^\mu=0$$

and these equations are completely integrable if

$$R_{\mu\nu,\alpha\beta}=K(g_{\nu\alpha}g_{\mu\beta}-g_{\mu\alpha}g_{\nu\beta}),$$

that is, for a space of constant curvature. Galilean space corresponds to vanishing curvature

$$R_{\mu\nu,\alpha\beta}=0.$$

What I wish to stress is that the properties of the uniform Galilean space-time can be expressed in a generally covariant manner. On the other hand, the Einsteinian gravitation theory supposes the space-time to be nonuniform. It is just this fundamental assumption, and not the general covariance of equations, that distinguishes the gravitation theory from the theory of the Galilean space-time.

This distinction has not been sufficiently understood, or in any case not sufficiently stressed, by many physicists and, paradoxical as it may seem, by Einstein himself, although the French mathematician Cartan has drawn attention to it many years ago. Einstein called both theories relativity theories. But what is relativity? This word has been misused. It is natural to connect the notion relativity with uniformity of space and time. The **uniformity** of Galilean space with respect to positions, directions, and nonaccelerated motions may be as well termed as **relativity** of positions, of directions, and of nonaccelerated motions. That is the true content of Einstein's principle of relativity of 1905. Use of the word relativity in this sense is quite legitimate.

But, if one does this, if one connects relativity with uniformity, the relativity has nothing to do with general covariance, that is, with covariance in which (1) is true but (2) is not necessarily satisfied. This means also that, **in the theory of nonuniform space-time, there is no principle of relativity.** The generalization of the theory which consists in replacement of time by a nonuniform one means a **restriction** and not a generalization of relativity. If one uses the word relativity consistently, then the general principle of relativity is nonsensical.

In saying this, I do not want to introduce any doubt as to the validity of the wonderful Einsteinian gravitation theory, but only to stress the inconsistency of the use of the name "general relativity" when applied to gravitation theory.

Einstein himself proposed for his theory the name "general relativity," because the transformations considered in this theory are more general than the Lorentz transformations. But he omitted to state that, in the case of ordinary relativity, one has to consider transformations for which (2) must also be fulfilled, while, in the case of the so-called general relativity, this equation does not have to be taken into account. Thus, in the "general" theory, Einstein uses the word relativity simply as covariance, while in the "special" theory, the same word relativity is used as uniformity. Since covariance has nothing to do with uniformity, there arises a confusion which is very harmful to the understanding of Einstein's theory. If one uses the word relativity in both senses, then one has to admit such statements as "in general relativity there is no relativity" or "the Lagrangian form of nonrelativistic equations of motion satisfies the requirements of general relativity," etc.

This confusion is more harmful than it would seem at first glance. It leads to statements like "rotation is relative" which are obviously false, because the distinction between a geodetic and a nongeodetic is absolute and not relative.

The general covariance of equations has been considered for a long time as a specific property of the Einsteinian gravitation theory, by which it is distinguished from other physical theories. But later on, it was recognized that the covariance by itself cannot lead to any physical consequences. The true key to Einstein's discovery and the most difficult step was the limitation of the functions describing the gravitational field to geometrical ones (to the $g_{\mu\nu}$'s). Historically, the covariance requirement played a great part also, but this is because it was combined with other requirements, such as simplicity and beauty of the theory.

Nevertheless, the covariance requirement is still considered in a somewhat mystical way, as something prohibiting the use of well-defined coordinate frames, like Galilean coordinates in uniform space-time. The existence of Galilean coordinate frames is a characteristic of the inherent properties of the uniform space-time of the "special" theory. Likewise, there may be in "general" theory coordinate frames distinguished by some remarkable properties and characteristic of the kind of the nonuniform space-time considered.

In what follows, I wish to draw your attention to the fact that, for a rather general class of problems of gravitation theory, there exist such coordinate systems that may be considered as generalizations of ordinary internal systems. I mean not the local geodetic system valid in the vicinity of a point and of an instant of time, but the nonlocal generalization of the inertial frames of reference, valid throughout space.

In order to investigate whether such systems exist, it is necessary to make definite assumptions as to the physical system considered and as to the properties of space-time as a whole. This is necessary because of the nonlocal character of the problem, that requires a solution of Einstein's gravitational equations with conditions at infinity.

In the case of an isolated system of masses, it is natural to consider the system as embedded in a Galilean space-time. In a Galilean space-time, the following *theorem* holds.

Let

$$\Box\langle\psi\rangle=\frac{1}{c^2}\frac{\partial^2\langle\psi\rangle}{\partial t^2}-\left(\frac{\partial^2\langle\psi\rangle}{\partial x^2}+\frac{\partial^2\langle\psi\rangle}{\partial y^2}+\frac{\partial^2\langle\psi\rangle}{\partial z^2}\right).$$

If ψ satisfies the wave equation $\Box\psi=0$ and is finite everywhere and tends to zero at infinity like $1/r$, as well as its derivatives, and if in addition the radiation condition

$$\lim_{r\to\infty}\left\{\frac{\partial(r\psi)}{\partial r}+\frac{1}{c}\frac{\partial(r\psi)}{\partial t}\right\}=0$$

is satisfied for all values of the time t, *then ψ vanishes identically.* The radiation condition states that only outgoing waves are allowed.

A similar theorem may be proved in the case that $\Box\psi$ refers to a static Einsteinian space-time which is Galilean at infinity. It is to be supposed that the theorem holds also for a nonstatic Einsteinian space-time, though a formal proof may be difficult.

Let the space-time be such that the aforestated theorem is valid. Then, one can introduce auxiliary conditions for the coordinates in such a way that they behave like Galilean coordinates and are determined like them throughout the space-time (a Lorentz transformation remains of course arbitrary).

The auxiliary conditions are of the form $\Box x_\nu = 0$; $\nu = 0, 1, 2, 3$. But we have

$$\Box\psi = g^{\mu\nu}\frac{\partial^2\psi}{\partial x_\mu \partial x_\nu} - \Gamma^\nu \frac{\partial\psi}{\partial x_\nu},$$

where

$$\Gamma^\nu = -\frac{1}{(-g)^{\frac{1}{2}}}\frac{\partial \mathfrak{g}^{\mu\nu}}{\partial x_\mu}; \quad \mathfrak{g}^{\mu\nu} = (-g)^{\frac{1}{2}} g^{\mu\nu}.$$

Consequently, the condition is equivalent to

$$\partial \mathfrak{g}^{\mu\nu}/\partial x_\mu = 0.$$

Let the coordinates x_μ satisfy this condition. To find the most general form

$$x_\alpha' = f^\alpha(x_0 x_1 x_2 x_3)$$

we put

$$f^\alpha = a_\alpha + l_\beta a_{\alpha\beta} x_\beta + \eta^\alpha. \quad (l_0 = 1;\, l_1 = l_2 = l_3 = -1)$$

The linear part of this is a Lorentz transformation. Now, η^α must satisfy the wave equation $\Box\eta^\alpha = 0$, since f^α and the linear part satisfy it. Further, η^α must vanish at infinity (because the transformation must reduce there to a Lorentz transformation) and also η^α must satisfy the radiation condition (this follows from the radiation condition for the $g^{\mu\nu}$'s):

$$\eta^\alpha = \text{outgoing wave at infinity.}$$

But the conditions imposed upon η^α are so stringent that, according to the theorem, $\eta^\alpha \equiv$ zero everywhere. Thus, the whole arbitrariness of the coordinates resides in the Lorentz transformation.

We thus come to the conclusion that, in the case of an isolated system of masses, there is no essential difference in the coordinate question, between the so-called general and so-called special relativity theory. In both cases, arbitrary coordinates are admissible, since the equations are, or may be, written covariantly with respect to general transformations. But, in both cases, auxiliary conditions may be imposed upon the coordinates in such a way that only a Lorentz transformation remains arbitrary.

The coordinates so defined—I call them harmonic —are particularly adapted to the solution of Einstein's equations, and all the solutions that I shall discuss in the following lectures are obtained in these coordinates. But the value of the harmonic coordinates resides not only in their practical importance, but also in the fact that they help us to understand the general features of gravitation theory. Their existence shows that the usual sharp distinction between the coordinate problem in special and in general theory is somewhat artificial. In both theories, coordinates exist that are determined to a Lorentz transformation, but in both theories any other coordinate system may be used.

SECOND LECTURE: SOME APPROXIMATE SOLUTIONS OF EINSTEIN'S EQUATIONS (MOTION OF ROTATING BODIES OF FINITE SIZE)

The first approach to the problem of the motion of finite masses in connection with the solution of Einstein's nonlinear field equations was made in 1927 in a paper by Einstein and Grommer, and another paper by Einstein. More definite progress was obtained, independently and on two different lines, in the years 1938–1940 by Einstein, Infeld, and their collaborators, on the one hand, and by myself and my pupils, Petrova, Fichtenholz, and others, on the other hand. I should also like here to mention the work by Papapetrou from 1951. The ideas underlying the two research lines were widely different. Einstein's intention was, as it seems, to build a theory of elementary particles as field singularities. My aim was quite different. I wanted to find the solution of Einstein's gravitational equations for a problem of astronomical kind, where the moving bodies are of finite size, and the field has no singularities even within the bodies. I first found an approximate solution for the case of spherical nonrotating bodies, and then for the more general case of rotating bodies that are not necessarily spherical. The equations of motion for the centers of spherical nonrotating bodies turned out to be the same as the equations for point singularities obtained at the same time by Einstein, Infeld, and others. But this coincidence is due, to some extent, to good luck, because of the coordinate problem. My solution is written in a well-defined coordinate system (the harmonic one), while Einstein's and Infeld's solution corresponds to some vaguely defined coordinate system which bears a resemblance to the Newtonian one, but is modified from step to step (from approximation to approximation). Professor Infeld sees even a virtue in the fact that the coordinate system he uses is not well defined, but remains to some extent arbitrary—a standpoint with which I cannot agree.

The general problem is to find the solution to Einstein's gravitational equations, which corresponds to the motion of a given system of masses. It is clear from the start that, in order to reduce this problem to a mechanical one, it is necessary to make approximations. First of all, because of the gravitation waves, the physical system considered is not a conservative one. But, since the loss of energy by gravitational radiation is quite small, it is safe to neglect it and to consider the system as conservative. Then, the bodies themselves have in general an infinite number of degrees of freedom even in the nonrelativistic Newtonian approximation. Some important conclusions may however be drawn quite generally: thus, one may write down the integrals of the equations of motion and the

asymptotic expressions for the fundamental tensor without reducing the inner degrees of freedom of the bodies to a finite number. But, in order to consider the motion of bodies as a whole, it is necessary to express the state of motion of each body in terms of a finite number of parameters.

There are thus two aspects of the problem of motion: the inner problem and the outer problem. In the Newtonian approximation, the conditions inside the bodies have no influence on the field outside them: only the total mass and the moments of inertia are important. The first relativistic approximations for the field outside requires a more detailed description of the motion inside the body, namely a description by means of Newtonian equations of motion for the continuous medium of which the body is built. This is natural, since in the first relativistic approximation one has to take into account the mass corresponding to the inner energy, and the latter may be calculated in the Newtonian approximation.

If, on the contrary, one considers the bodies as field singularities, the idealization is so far-reaching that the inner problem does not arise at all.

The formal procedure in solving Einstein's equation is to develop the solution in powers of q/c, q being a parameter of the order of the velocities involved. We have then, U being the Newtonian potential,

$$U/c^2 \sim q^2/c^2 \quad \text{and} \quad v^2/c^2 \sim q^2/c^2.$$

Since retardation is treated as a correction, the method is valid only for moderately large distances between and from the bodies, that is, for distances small compared with the wavelength emitted. If R is a distance and ω a typical frequency, we must have

$$R \ll c/\omega.$$

This condition is not independent of the condition $q \sim c$, since q and $R\omega$ are of the same order.

On the other hand, in order to get convenient expressions for the potential from the bodies, we must suppose the distances R to be large compared with the dimensions L of the bodies. We thus have

$$L \ll R \ll c/\omega.$$

Inside the body, the quantity U/c^2 may attain the value α/L, where α is the gravitational radius of the body. Since this is to be small, we have $\alpha \ll L$, and thus

$$\alpha \ll L \ll R \ll c/\omega.$$

To solve Einstein's equations, step by step, we proceed as follows. In the zeroth approximation, we consider the metrics as Euclidean, and the coordinates as Galilean, and we take for the energy tensor the expressions,

$$T^{00} = \frac{1}{c^2}\rho, \quad T^{0i} = \frac{1}{c^2}\rho v_i \quad (i=1,2,3),$$

(ρ density), while the values of the other components T^{ik} are irrelevant [they are of the order $T^{ik} \sim \rho(q^2/c^2)$]. The foregoing expressions satisfy the relation

$$\frac{\partial T^{00}}{\partial t} + \frac{\partial T^{0i}}{\partial x_i} = 0$$

since

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v_i)}{\partial x_i} = 0$$

(it is summed from 1 to 3 over Latin indexes appearing twice). To make the next step, we calculate the Newtonian potential

$$U = \int \frac{\gamma \rho'}{|\mathbf{r}-\mathbf{r}'|} dV'$$

and the corresponding vector potential

$$U_i = \int \frac{\gamma(\rho v_i)'}{|\mathbf{r}-\mathbf{r}'|} dV'.$$

This gives for the metric tensor the approximation

$$g_{00} = c^2 - 2U; \quad g_{0i} = \frac{4}{c^2} U_i$$

$$g_{ik} = -\left(1 + \frac{2U}{c^2}\right)\delta_{ik}$$

from which the quantities $\mathfrak{g}^{\mu\nu} = (-g)^{\frac{1}{2}} g^{\mu\nu}$ are easily calculated, namely

$$\mathfrak{g}^{00} = \frac{1}{c} + \frac{4U}{c^3}; \quad \mathfrak{g}^{0i} = \frac{4U_i}{c^3}; \quad \mathfrak{g}^{ik} = -c\delta_{ik}.$$

These quantities satisfy the equation,

$$\frac{\partial \mathfrak{g}^{00}}{\partial t} + \frac{\partial \mathfrak{g}^{0i}}{\partial x_i} = 0.$$

In the approximation considered, the coordinates are thus harmonic.

Proceeding to the next step, we must take into account the corrections in the expression for the divergence of a tensor. We have now

$$\frac{\partial T^{00}}{\partial t} + \frac{\partial T^{0k}}{\partial x_k} + \frac{1}{c^2}\frac{\partial U}{\partial t} T^{00} = 0$$

and

$$\frac{\partial T^{0i}}{\partial t} + \frac{\partial T^{ik}}{\partial x_k} - \frac{\partial U}{\partial x_i} T^{00} = 0.$$

Supposing the bodies to be elastic, we may take for the energy tensor the expressions

$$c^2T^{00}=\rho\left\{1+\frac{1}{c^2}(\tfrac{1}{2}v^2+\Pi-U)\right\}$$

$$c^2T^{0i}=\rho v_i\left\{1+\frac{1}{c^2}(\tfrac{1}{2}v^2+\Pi-U)\right\}-\frac{1}{c^2}p_{ik}v_k$$

$$c^2T^{ik}=\rho v_iv_k-p_{ik}.$$

These expressions are obtained by adding to that part of mass density, which is conserved, the variable part of the mass density and also by adding terms corresponding to the energy flux. The notion of energy flux was first introduced by the Russian physicist Umow as early as in 1874 so that the expressions above are essentially the same as in the Umow paper.

The quantities ρ etc. entering into these expressions satisfy the nonrelativistic equations of motion, namely

$$\rho\frac{dv_i}{dt}-\rho\frac{\partial U}{\partial x_i}=\frac{\partial p_{ik}}{\partial x_k}$$

as well as the continuity equation

$$\frac{\partial\rho}{\partial t}=\frac{\partial(\rho v_i)}{\partial x_i}=0.$$

The letter Π denotes the elastic energy density that satisfies the equation

$$\rho\frac{d\Pi}{dt}=\tfrac{1}{2}p_{ik}\left(\frac{\partial v_i}{\partial x_k}+\frac{\partial v_k}{\partial x_i}\right).$$

In virtue of these equations, it is easily proved that the divergence of the energy tensor vanishes in the required approximation.

Now, the calculations can proceed in different directions. Firstly, one may obtain, from the Einsteinian equations

$$R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R=-\frac{8\pi\gamma}{c^2}T^{\mu\nu}$$

the next approximation for the $\mathfrak{g}^{\mu\nu}$'s. Secondly, one may obtain the equations of motion for the bodies without solving Einstein's equations explicitly. One can also write down the integrals of the equations of motion. Since the calculations are rather tedious, I shall only briefly indicate their principal idea and the results.

In a harmonic coordinate system, where the condition

$$\partial\mathfrak{g}^{\mu\nu}/\partial x_\mu=0$$

is satisfied, we have

$$g(R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R)=\tfrac{1}{2}\mathfrak{g}^{\alpha\beta}\frac{\partial^2\mathfrak{g}^{\mu\nu}}{\partial x_\alpha\partial x_\beta}+\cdots,$$

where the members denoted by dots do not contain second derivatives. If one uses this expression, the vanishing of divergence of the tensor $T^{\mu\nu}$ is no longer an identity, but is only satisfied in virtue of the harmonicity condition. Approximately, we have

$$\left(\Delta-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\frac{\partial\mathfrak{g}^{\mu\nu}}{\partial x_\mu}=\frac{16\pi\gamma}{c^3}g\nabla_\mu T^{\mu\nu},$$

where Δ is the Euclidean operator. In order that $\partial\mathfrak{g}^{\mu\nu}/\partial x_\mu$ should decrease rapidly with the distance from the mass m_a, it is necessary that

$$\int_{(a)} g\nabla_\mu T^{\mu\nu}(dx)^3=0;\quad (dx)^3=dx_1dx_2dx_3$$

where the integral is taken over a volume enclosing the mass m_a. This gives the equations of motion for the centers of inertia of the separate bodies. To obtain equations for rotational motion, one has to write

$$\int_{(a)} g(x_i\nabla_\mu T^{\mu k}-x_k\nabla_\mu T^{\mu i})(dx)^3=0.$$

I shall not write down the equations of motion in explicit form because they are complicated. But it is worth noting that the equations of motion for the centers of mass can be written in Lagrangian form.

From the equations of motion, the ten classical integrals may be obtained (namely, the momentum and energy integral, the moment of momentum, and the center of gravity integral). These integrals may be deduced by integrating the previous expressions over the whole space instead of over the region where a single mass lies.

The integrals have the following form. Let

$$W=\tfrac{1}{2}\gamma\int\rho'|\mathbf{r}-\mathbf{r}'|(dx')^3$$

so that

$$\Delta W=U=\gamma\int\frac{\rho'}{|\mathbf{r}-\mathbf{r}'|}(dx')^3,$$

where U is the Newtonian potential, and let U_i be, as before, the vector potential.

We put

$$G_i=\left[\rho+\frac{\rho}{c^2}(\tfrac{1}{2}v^2+\Pi+3U)\right]v_i-\frac{1}{c^2}p_{ij}v_j-\frac{4}{c^2}\rho U_i-\frac{1}{c^2}\rho\frac{\partial^2W}{\partial x_i\partial t}.$$

Then, the quantity,

$$P_i=\int G_i(dx)^3,$$

which is to be interpreted as the momentum of the system, is constant in virtue of the equations of motion.

If we put

$$M_{ik}=\int (x_i G_k - x_k G_i)(dx)^3,$$

then M_{ik} is the angular momentum and is also constant.

The energy integral is

$$M=\int \rho\left\{1+\frac{1}{c^2}\left(\tfrac{1}{2}v^2+\Pi-\tfrac{1}{2}U\right)\right\}(dx)^3.$$

If one uses the Lagrangian form of the equations of motion one can obtain the next term (with respect to q^2/c^2) in the energy integral also.

Finally, for the integral of the center of mass, we have

$$MX_i - P_i t = K_i,$$

where the center of mass is defined by

$$MX_i=\int x_i\rho\left\{1+\frac{1}{c^2}\left(\tfrac{1}{2}v^2+\Pi-\tfrac{1}{2}U\right)\right\}(dx)^3.$$

All the quantities M, P_i, M_{ik}, K_i are constant in virtue of the relativistic equations of motion.

The values of these constants are essential for the asymptotic behavior of the metric tensor at large distances from the system of masses. We write these expressions for moderately large distances (not in the wave region). For the space components $\mathfrak{g}^{ik}$, we have

$$\mathfrak{g}^{ik}=-c\delta_{ik}+\frac{2\gamma}{c^3 r}\frac{d^2}{dt^2}\int \rho x_i x_k (dx)^3$$
$$+\frac{2\gamma}{c^3}\frac{x_j}{r^3}\frac{d}{dt}\int \rho(x_j x_i v_k + x_j x_k v_i - x_i x_k v_j)(dx)^3+\frac{\gamma^2 M^2}{c^3 r^4}x_i x_k.$$

Introducing the generalized moments of inertia

$$D_{ji}=\int \rho x_i x_j\left\{1+\frac{1}{c^2}\left(\tfrac{1}{2}v^2+\Pi-\tfrac{1}{2}U\right)\right\}(dx)^3$$
$$+\frac{7}{2c^3}\delta_{ij}\int \rho W(dx)^3$$

and also the function

$$W_i=\tfrac{1}{2}\gamma\int (\rho v_i)'\,|\mathbf{r}-\mathbf{r}'|\,(dx')^3$$

which allows us to take into account the retardation in the vector potential, then we obtain

$$\mathfrak{g}^{0i}=\frac{4\gamma}{c^2 r}P_i+\frac{2\gamma x_j}{c^3 r^3}M_{ji}-\frac{\partial^2}{\partial x_j \partial t}\frac{2\gamma}{c^3 r}D_{ji}+\frac{\partial^2}{\partial x_j \partial x_k}\left(\frac{2\gamma}{c^3 r}\right)$$
$$\times\int \rho v_i x_j x_k (dx)^3+\frac{7\gamma^2 M P_i}{c^5 r^2}+\frac{\gamma^2 M P_k x_i x_k}{c^5 r^4}+\frac{4}{c^5}\frac{\partial^2 W_i}{\partial t^2}.$$

The last three terms are correction terms of the order q^2/c^2 with respect to the main terms.

For the component $\mathfrak{g}^{00}$ of the fundamental tensor, we obtain finally

$$\mathfrak{g}^{00}=\frac{1}{c}+\frac{4\gamma M}{c^3 r}+\frac{4\gamma x_j}{c^3 r^3}MX_j+\frac{\partial^2}{\partial x_j \partial x_k}\left(\frac{2\gamma}{c^3 r}\right)D_{jk}$$
$$-\frac{\partial^3}{\partial x_i \partial x_j \partial x_k}\left(\frac{2\gamma}{c^3 r}\right)\int \rho x_i x_j x_k (dx)^3+\frac{7\gamma^2 M^2}{c^5 r^2}$$
$$+\frac{14\gamma^2 M^2 X_j x_j}{c^5 r^4}+\frac{4}{c^5}\frac{\partial^2 W}{\partial t^2}.$$

All the expressions are of course approximate. It is easy to see that $\mathfrak{g}^{00}$ and $\mathfrak{g}^{0i}$ verify the relation

$$\frac{\partial \mathfrak{g}^{00}}{\partial t}+\frac{\partial \mathfrak{g}^{0i}}{\partial x_i}=0.$$

The corresponding relation for $\mathfrak{g}^{0i}$ and $\mathfrak{g}^{ik}$ is also verified up to terms containing c^3 in the denominator (terms of higher order have been neglected in $\mathfrak{g}^{ik}$). Thus, the coordinates used are in fact harmonic.

In the wave region, the foregoing expressions are to be modified. I consider this modification in the next lecture.

THIRD LECTURE: ON GRAVITATIONAL WAVES FROM A SYSTEM OF MOVING BODIES

In my previous lecture, I have given asymptotic expressions for the metric tensor, valid at moderately large distances from the system of moving bodies. (The distances must be large compared with dimensions of the system, but still small compared with the wavelength of the gravitational waves emitted.) There are questions, however, for which the knowledge of the asymptotic values of the metric tensor at very large distances is required, that is, at distances of the order of the wavelength, or even large, compared with the wavelength. These questions are of a theoretical nature, but nevertheless they are of interest, because, the gravitational equations being nonlinear, it is not so simple to establish even the existence of gravitational waves.

In the following, I intend to show that there exist solutions corresponding to spherical waves emitted by the system of moving bodies. This is due to the particular structure of gravitational equations in which the components $\mathfrak{g}^{\mu\nu}$ of the metric tensor are at the same time coefficients in the wave operator and unknown functions to which this operator applies.

Einstein's equation in harmonic coordinates may be written in the form

$$g\left(R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R\right)=\tfrac{1}{2}\mathfrak{g}^{\alpha\beta}\frac{\partial^2 \mathfrak{g}^{\mu\nu}}{\partial x_\alpha \partial x_\beta}-c^2 N^{\mu\nu}=-\frac{8\pi\gamma g}{c^2}T^{\mu\nu},$$

where $N^{\mu\nu}$ is a quadratic form in the first derivatives of the $\mathfrak{g}^{\alpha\beta}$'s. As the zeroth approximation to the wave operator we choose

$$\Box^0\psi=\frac{1}{c^2}\left(1+\frac{4\alpha}{r}\right)\frac{\partial^2\psi}{\partial t^2}-\left(\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}+\frac{\Delta^*\psi}{r^2}\right)$$

where

$$\Delta^*\psi=\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial\psi}{\partial\vartheta}\right)+\frac{1}{\sin^2\vartheta}\frac{\partial^2\psi}{\partial\varphi^2}$$

is the angular part of the ordinary Laplace operator in spherical coordinates. In the coefficients of the operator $\Box^0$ only statical terms of the order $1/r$ are retained.

We want to study spherical waves going out from the system. For such waves, the term $\Delta^*\psi/r^2$ in the wave operator is small as compared with other terms and we can replace the wave equation by

$$\frac{1}{c^2}\left(1+\frac{4\alpha}{r}\right)\frac{\partial^2\psi}{\partial t^2}-\left(\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}\right)=0.$$

If we write

$$r\psi=f$$

and introduce the independent variable

$$r^*=r+2\alpha(\lg r-\lg r_0)$$

we have approximately

$$\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}-\frac{\partial^2 f}{\partial r^{*2}}=0.$$

The solution that corresponds to the outgoing wave is of the form

$$f=f(\tau,\mathbf{n}),$$

where $\mathbf{n}$ is the unit vector in the direction *from* the system, with components $n_i=(x_i/r)$ and τ is the quantity

$$\tau=t-\frac{1}{c}r^*=t-\frac{1}{c}\left(r+2\alpha\lg\frac{r}{z_0}\right).$$

Thus, the outgoing wave has the asymptotic form

$$\psi=\frac{1}{r}f(\tau,\mathbf{n}).$$

For a function of this form, derivatives with respect to the coordinates and the time can be calculated by neglecting the dependence of ψ on $\mathbf{n}$ and r and by taking into account only the dependence on τ.

Putting

$$k_\alpha=\partial\tau/\partial x_\alpha,$$

and thus

$$k_0=1;\quad k_i\cong-n_i/c,$$

we have approximately

$$\partial\psi/\partial x_\alpha=k_\alpha\dot\psi, \tag{*}$$

where the point indicates the derivatives with respect to τ (or to t).

In the above expression for $\Box^0\psi$ only the static part of $\mathfrak{g}^{\mu\nu}$ has been taken into account, but the non-static part (which we denote by $C^{\mu\nu}$) may be of the same order of magnitude. We may put

$$\mathfrak{g}^{00}=\frac{1}{c}+\frac{4\alpha}{cr}+C^{00}$$

$$\mathfrak{g}^{0i}=C^{0i}$$

$$\mathfrak{g}^{ik}=-c\delta_{ik}+C^{ik}$$

where the $C^{\mu\nu}$'s are of the same asymptotic form as ψ, so that

$$\partial C^{\mu\nu}/\partial x_\alpha=k_\alpha\dot C^{\mu\nu}.$$

But the derivatives of the static part of the $\mathfrak{g}^{\mu\nu}$'s decrease as $1/r^2$ and are to be neglected as compared with the derivatives of the $C^{\mu\nu}$'s. Thus, we may write

$$\partial\mathfrak{g}^{\mu\nu}/\partial x_\alpha=k_\alpha\dot C^{\mu\nu}. \tag{**}$$

Now, the harmonicity condition gives

$$\partial\mathfrak{g}^{\mu\nu}/\partial x_\mu=k_\mu\dot C^{\mu\nu}=0$$

and this may be integrated (with respect to τ or to t). We thus obtain

$$k_\mu C^{\mu\nu}=0.$$

The integration constant has been put equal to zero, because the the static part of the $\mathfrak{g}^{\mu\nu}$'s is considered separately. From the last equation we have

$$C^{0i}=\frac{n_k}{c}C^{ik};\quad C^{00}=\frac{n_i n_k}{c^2}C^{ik}.$$

With the corrected values of $\mathfrak{g}^{\mu\nu}$, the expression for the wave operator becomes

$$\frac{1}{c}\mathfrak{g}^{\alpha\beta}\frac{\partial^2\psi}{\partial x_\alpha\partial x_\beta}=\Box^0\psi+\frac{1}{c}C^{\alpha\beta}\frac{\partial^2\psi}{\partial x_\alpha\partial x_\beta}.$$

But, if ψ is an outgoing wave and satisfies (*), then

$$C^{\alpha\beta}\frac{\partial^2\psi}{\partial x_\alpha\partial x_\beta}=k_\alpha k_\beta C^{\alpha\beta}\ddot\psi=0$$

separately. Thus, in the wave operator applied to an outgoing wave, the coefficients may be replaced by their static values. This justifies the asymptotic form of wave function used in the preceding calculations. (This form was obtained with the static values of the $\mathfrak{g}^{\mu\nu}$'s.)

We now proceed to the calculations of that part of the Einstein tensor that is quadratic in the first deriva-

tives of the $\mathfrak{g}^{\mu\nu}$. Using the expression (**) for $\partial\mathfrak{g}^{\mu\nu}/\partial x_\alpha$ and raising and lowering the indexes with Galilean values of the metric tensor we obtain

$$N^{\mu\nu}=\frac{1}{4c^2}k^\mu k^\nu(\dot{C}^{\alpha\beta}\dot{C}_{\alpha\beta}-\tfrac{1}{2}\dot{C}_\alpha{}^\alpha\dot{C}_\beta{}^\beta).$$

The quantity

$$\sigma_g=\frac{1}{32\pi\gamma}(\dot{C}^{\alpha\beta}\dot{C}_{\alpha\beta}-\tfrac{1}{2}\dot{C}_\alpha{}^\alpha\dot{C}_\beta{}^\beta)$$

is to be interpreted as the energy density of the gravitational waves. We have

$$N^{\mu\nu}=\frac{8\pi\gamma}{c^2}\sigma_g k^\mu k^\nu.$$

We may also take into account the electromagnetic radiation. Putting

$$\sigma_{\text{em}}=(E^2+H^2)/8\pi,$$

we have for the energy tensor in the wave zone

$$T^{\mu\nu}=\sigma_{\text{em}}k^\mu k^\nu.$$

Introducing the total density

$$\sigma=\sigma_g+\sigma_{\text{em}}$$

of radiation energy, we obtain the Einstein equations in the form

$$\frac{1}{2c}\Box\mathfrak{g}^{\mu\nu}=\frac{8\pi\gamma}{c^2}\sigma k^\mu k^\nu.$$

From this, it is clear that the correction to the $\mathfrak{g}^{\mu\nu}$'s due to the radiation terms in Einstein's equations (to the terms $N^{\mu\nu}$ and $T^{\mu\nu}$) is of the form $hk^\mu k^\nu$, where h satisfies the equation

$$\Box h=\frac{16\pi\gamma}{c}\sigma.$$

This correction modifies by the same amount the values of $C^{\mu\nu}$ and it might seem that this modification alters the value of the gravitational energy density σ_g. If it were so, then the integration problem would not be solved and should be taken up anew. As a matter of fact, it happens, however, that, owing to the relation $k_\mu\dot{C}^{\mu\nu}=0$, the expression for σ_g is *invariant* with respect to the transformation

$$\dot{C}^{\alpha\beta}\rightarrow\dot{C}^{\alpha\beta}+\lambda k^\alpha k^\beta.$$

Thus, the right-hand side of the equation for $\mathfrak{g}^{\mu\nu}$ is not to be modified.

The radiation energy density σ is of the form

$$\sigma=\frac{\sigma_0(\tau,\mathbf{n})}{r^2}.$$

Inserting this in the equation for h, we find the following asymptotic expression for this quantity

$$h=\frac{8\pi\gamma\lg r}{r}\int_{\tau_0}^{\tau}\sigma_0(\tau,\mathbf{n})d\tau+\frac{h_0(\tau,\mathbf{n})}{r}.$$

This may be written as

$$h=\frac{2\gamma}{cr}(\lg r\,\Delta E+\epsilon),$$

where $\Delta E(d\Omega/n\pi)$ is the energy loss during the time $\tau-\tau_0$ in the solid angle $d\Omega$ lying in the direction $\mathbf{n}$. Superposing a solution of the homogeneous wave equation, we may write

$$C^{\mu\nu}=\frac{2\gamma}{c^3 r}f^{\mu\nu}(\tau,\mathbf{n})+\frac{2\gamma}{cr}(\lg r\Delta E+\epsilon)k^\mu k^\nu.$$

We thus have in the wave region

$$\mathfrak{g}^{00}=\frac{1}{c}+\frac{4\gamma M}{c^3 r}+\frac{2\gamma}{c^3 r}n_i n_k f^{ik}+\frac{2\gamma}{c^3 r}(\lg r\,\Delta E+\epsilon)$$

$$\mathfrak{g}^{0i}=\frac{2\gamma}{c^4 r}n_k f^{ik}+\frac{2\gamma}{c^4 r}n_i(\lg r\,\Delta E+\epsilon)$$

$$\mathfrak{g}^{ik}=-c\delta_{ik}+\frac{2\gamma}{c^3 r}f^{ik}+\frac{2\gamma}{c^3 r}n_i n_k(\lg r\,\Delta E+\epsilon)$$

since

$$f^{0i}=\frac{n_k}{c}f^{ik};\quad f^{00}=\frac{n_i n_k}{c^2}f^{ik}.$$

We now may compare the expressions just obtained with the former expressions valid at moderately large distances from the system.

If we introduce the moments of inertia

$$D_{ik}(t)=\int\rho x_i x_k(dx)^3$$

and replace t by τ, then we can put

$$f^{ik}=(d^2/d\tau^2)D_{ik}(\tau).$$

Neglecting terms in ΔE and ϵ that are very small (in spite of the logarithm), we find that the formulas

$$\mathfrak{g}^{00}=\frac{1}{c}+\frac{4\gamma M}{c^3 r}+\frac{2\gamma}{c^3}\frac{\partial^2}{\partial x_i\partial x_k}\frac{D_{ik}(\tau)}{r}$$

$$\mathfrak{g}^{0i}=-\frac{2\gamma}{c^3}\frac{\partial^2}{\partial x_k\partial t}\frac{D_{ik}(\tau)}{r}$$

$$\mathfrak{g}^{ik}=-c\delta_{ik}+\frac{2\gamma}{c^3}\frac{\partial^2}{\partial t^2}\frac{D_{ik}(\tau)}{r}$$

give approximately correct values for the $\mathfrak{g}^{\mu\nu}$'s in the wave zone as well as at moderate distances from the system.

The harmonicity condition is satisfied rigorously.

I should like to add some remarks as to the energy pseudo-tensor and the general formulation of conservation laws.

In virtue of Einstein's equations, one can introduce the following symmetrical pseudo-tensor

$$U^{\mu\nu}=\frac{1}{16\pi\gamma}\frac{\partial^2}{\partial x_\alpha \partial x_\beta}(\mathfrak{g}^{\alpha\beta}\mathfrak{g}^{\mu\nu}-\mathfrak{g}^{\alpha\mu}\mathfrak{g}^{\beta\nu}).$$

This pseudo-tensor is symmetrical and satisfies the equation,

$$\partial U^{\mu\nu}/\partial x_\mu=0,$$

identically.

Now consider the integrals

$$*M=c^2\int U^{00}(dx)^3$$

$$*P^i=c^2\int U^{0i}(dx)^3$$

$$*M^{ik}=c^2\int (x_i U^{0k}-x_\mu U^{0i})(dx)^3$$

$$*M^{i0}=c^2\int (x_i U^{00}-tU^{0i})(dx)^3$$

$$=*M*X^i-t*P^i$$

taken over a finite volume V. The time derivatives of these integrals may be transformed into surface integrals. The surface enclosing the volume V (which contains all the masses) is to be taken at moderately large distances from the system of masses. Then the surface integrals will be very small and thus the volume integrals nearly constant.

But the constancy of the volume integrals may be verified by direct calculation. The integrals themselves, and not only their time derivatives, may be transformed into surface integrals. If they are taken over a distant surface, only asymptotic values of the $\mathfrak{g}^{\mu\nu}$'s are to be used. Inserting these in the integrals we obtain for the quantities $*M$, $*P^i$, $*M^{ik}$, $*M^{i0}$ just the constants M, P_i, M_{ik}', K_i entering in the asymptotic expressions, while all other terms in $\mathfrak{g}^{\mu\nu}$, like D_{ik}, etc., cancel out. We have thus a direct verification of the asymptotic expressions for the fundamental tensor.

As to the energy loss due to gravitational radiation, it can be calculated as usual with the help of the energy pseudo tensor which is approximately equal to

$$U^{\mu\nu}=\sigma_0 k^\mu k^\nu.$$

Introducing the quantities,

$$B_{ik}=\frac{d^3}{d\tau^3}\{D_{ik}(\tau)-\tfrac{1}{3}\delta_{ik}D_{jj}(\tau)\},$$

we obtain, for the rate of energy loss,

$$\frac{dW}{dt}=-\frac{\gamma}{5c^5}B_{ik}B_{ik},$$

which is an extremely small quantity owing to the very large value of the constant

$$5c^3/\gamma=2.10^{39}\text{ gram/sec.}$$

In these three lectures, I have given only a very brief account of the problems in relativity theory that have interested me during recent years. A more detailed treatment of these problems, with all the calculations involved, is contained in my book *Theory of Space, Time, and Gravitation* published in Russian in 1955, and which will appear in English translation in 1957 or 1958.

Einstein's and Other Theories of Gravitation*

SURAJ N. GUPTA

Department of Physics, Wayne State University, Detroit, Michigan

I. INTRODUCTION

EVER since it has been realized that Newton's theory of gravitation has its limitations, several theories of gravitation have been proposed from time to time. The aim of this paper is to discuss the various theories of gravitation with particular attention to Einstein's theory. We consider only the pure gravitational theories, and thus do not discuss attempts to find a unified field theory of various fields in nature.

If we leave aside all philosophical considerations, it seems reasonable that any acceptable theory of gravitation should satisfy the following requirements:

(1) It should be Lorentz covariant, because the special theory of relativity has now been well established by experiments.

(2) It should reduce to Newton's theory of gravitation as a good approximation, because Newton's theory is able to explain the observed gravitational phenomena to a fairly high degree of accuracy.

(3) It should also provide a reasonable explanation for the so-called three crucial tests.

Keeping the above requirements in mind, we first consider some attempts to construct a theory of the gravitational field in flat space similar to the theories of the electromagnetic field and the meson fields. We then show that Einstein's theory itself can be regarded as a theory of gravitation in flat space, and describe the advantages of such an approach from logical as well as practical points of view.

In this paper we deal with tensors in flat space as well as in the Riemannian space. In flat space the Greek indices will take the values 1, 2, 3, 4, and the space-time coordinates will be denoted as $x_\mu=(x_1,x_2,x_3,ict)$. In the Riemannian space the Greek indices will take the values 1, 2, 3, 0, and the space-time coordinates will be denoted as $x^\mu=(x^1,x^2,x^3,ct)$.

II. LORENTZ-COVARIANT THEORIES OF GRAVITATION

We follow the usual ideas of the field theory to see whether it is possible to find a theory of gravitation in flat space. In order that our theory of gravitation may reduce to Newton's theory as a good approximation, our gravitational field on quantization should correspond to neutral particles of vanishing rest mass and integral spin. Hence, confining attention to fields of spin 0, 1, or 2, the possible field equations for the gravitational field are

$$\Box^2 U=\kappa T, \tag{1}$$

$$\Box^2 U_\mu=\kappa T_\mu, \tag{2}$$

$$\Box^2 U_{\mu\nu}=\kappa T_{\mu\nu}, \tag{3}$$

where the scalar U, the four-vector U_μ, and the symmetrical tensor $U_{\mu\nu}$ are real field variables; T, T_μ, and $T_{\mu\nu}$ are the source functions; and κ is the coupling constant. We do not here consider the more complicated field equations corresponding to particles of spin higher than 2.

The field equation (1) for the gravitational field was first discussed by Nordström,[1] and recently a very clear presentation of this theory has been given by Bergmann.[2] In this theory the source function T is the trace of the energy-momentum tensor of matter, which includes all particles and fields except the gravitational field. This theory satisfies the first two requirements of Sec. I, but for the advance of the perihelion of planets it gives a value, which is one-sixth of Einstein's value in magnitude and opposite in sign. Since this result is definitely contrary to experiments, this simple theory of the gravitational field of spin 0 is unacceptable.

The field equation (2) is of a very familiar form. It is also well known that, in order that the energy of such a field may be positive definite, U_μ must satisfy the supplementary condition

$$\partial U_\mu/\partial x_\mu=0, \tag{4}$$

whence it follows that the source function T_μ must also satisfy the relation

$$\partial T_\mu/\partial x_\mu=0. \tag{5}$$

The only known four-vector quantity, which satisfies (5), is the current four-vector. It is, therefore, evident that the gravitational field (2) will be identical with the electromagnetic field, except that the gravitational charge of a particle might be different from its electromagnetic charge. Such a theory of the gravitational field has to be rejected, because the observed properties of the gravitational field are quite different from those of the electromagnetic field. For instance, the gravitational force between any two particles is always attractive, while the electromagnetic force between like particles is repulsive. It is rather curious that this difficulty arises in the case of the field of spin 1, while the fields of spin 0 and 2 both lead to a gravitational interaction of the observed sign between any two particles.

* Supported in part by the National Science Foundation.

[1] G. Nordström, Ann. Physik 43, 1101 (1914).
[2] O. Bergmann, Am. J. Phys. 24, 38 (1956).

Birkhoff[3] has suggested an interesting theory of gravitation, which is based on a field equation of the form (3), where the source function $T_{\mu\nu}$ is the symmetrical energy-momentum tensor of matter. Birkhoff's theory satisfies the three requirements mentioned in Sec. I. Nevertheless, as pointed out by Weyl,[4] this theory suffers from serious difficulties. The essential difficulty here is that Birkhoff's gravitational field does not have a positive definite energy, and therefore on quantization it corresponds to particles of positive as well as negative energies. But particles of negative energy are not permissible in the quantum theory of fields, because otherwise the state of vacuum will become unstable due to the spontaneous production of real particles of positive and negative energies.

The above-mentioned difficulty in Birkhoff's theory arises from the fact that we cannot impose the supplementary condition

$$\partial U_{\mu\nu}/\partial x_\nu = 0 \tag{6}$$

on Birkhoff's gravitational field. For it would follow from (3) and (6) that

$$\partial T_{\mu\nu}/\partial x_\nu = 0, \tag{7}$$

which is not possible, because the energy-momentum tensor of matter alone cannot satisfy the conservation equation. In fact, instead of (7) we must have

$$\frac{\partial (T_{\mu\nu}+t_{\mu\nu})}{\partial x_\nu} = 0, \tag{8}$$

where $t_{\mu\nu}$ is the energy-momentum tensor of the gravitational field. This shows that the supplementary condition (6) will be compatible with the gravitational field equation, provided that we replace (3) by

$$\Box^2 U_{\mu\nu} = \kappa (T_{\mu\nu}+t_{\mu\nu}). \tag{9}$$

In the absence of matter, the above field equation reduces to

$$\Box^2 U_{\mu\nu} = \kappa t_{\mu\nu}, \tag{10}$$

which is a nonlinear equation for the gravitational field. Moreover, when we try to derive this field equation by the usual variational principle, we find[5] that the required Lagrangian density L has to be an infinite series of the form

$$L = L_0 + \kappa L_1 + \kappa^2 L_2 + \cdots + \kappa^n L_n + \cdots, \tag{11}$$

where the terms in L_n consist of a product of $(n+2)$ factors, each factor being $U_{\mu\nu}$ or its derivative. This peculiar situation in the case of the gravitational field of spin 2 is of particular interest in the next section.

III. EINSTEIN'S THEORY OF GRAVITATION

The most widely accepted theory of gravitation at the present time is due to Einstein. Einstein's theory not only satisfies the requirements of Sec. I, but it also possesses a beautiful mathematical structure. Usually Einstein's theory is regarded as a theory of gravitation in the Riemannian space, which makes this theory strikingly different from other field theories. We shall, however, show that Einstein's theory can also be treated as a theory of gravitation in flat space.

The Lagrangian density for Einstein's gravitational field in the Riemannian space is given by

$$\mathfrak{L} = -\kappa^{-2}\mathfrak{g}^{\mu\nu}\left(\begin{Bmatrix}\alpha\\ \mu\beta\end{Bmatrix}\begin{Bmatrix}\beta\\ \nu\alpha\end{Bmatrix} - \begin{Bmatrix}\alpha\\ \mu\nu\end{Bmatrix}\begin{Bmatrix}\beta\\ \alpha\beta\end{Bmatrix}\right), \tag{12}$$

where

$$\mathfrak{g}^{\mu\nu} = (-g)^{\frac{1}{2}} g^{\mu\nu}, \tag{13}$$

and

$$\begin{Bmatrix}\alpha\\ \mu\nu\end{Bmatrix} = \tfrac{1}{2} g^{\alpha\lambda}\left(\frac{\partial g_{\mu\lambda}}{\partial x^\nu} + \frac{\partial g_{\nu\lambda}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\right). \tag{14}$$

We now put

$$\mathfrak{g}^{\mu\nu} = \epsilon^{\mu\nu} - \kappa\gamma^{\mu\nu}, \tag{15}$$

where $\epsilon^{\mu\nu}$ is the Minkowskian metrical tensor

$$\epsilon^{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We can then express the tensors $g^{\mu\nu}$ and $g_{\mu\nu}$ as infinite series in powers of κ as[6]

$$\begin{aligned} g^{\mu\nu} = \epsilon^{\mu\nu} &+ \kappa(-\gamma^{\mu\nu} + \tfrac{1}{2}\epsilon^{\mu\nu}\epsilon_{\alpha\beta}\gamma^{\alpha\beta}) \\ &+ \kappa^2(-\tfrac{1}{2}\epsilon_{\alpha\beta}\gamma^{\alpha\beta}\gamma^{\mu\nu} + \tfrac{1}{4}\epsilon^{\mu\nu}\epsilon_{\alpha\rho}\epsilon_{\beta\lambda}\gamma^{\alpha\beta}\gamma^{\lambda\rho} \\ &+ \tfrac{1}{8}\epsilon^{\mu\nu}\epsilon_{\alpha\beta}\epsilon_{\lambda\rho}\gamma^{\alpha\beta}\gamma^{\lambda\rho}) + 0(\kappa^3), \end{aligned} \tag{16}$$

$$\begin{aligned} g_{\mu\nu} = \epsilon_{\mu\nu} &+ \kappa(-\tfrac{1}{2}\epsilon_{\mu\nu}\epsilon_{\alpha\beta}\gamma^{\alpha\beta} + \epsilon_{\mu\alpha}\epsilon_{\nu\beta}\gamma^{\alpha\beta}) \\ &+ \kappa^2(\epsilon_{\alpha\mu}\epsilon_{\lambda\nu}\epsilon_{\beta\rho}\gamma^{\alpha\beta}\gamma^{\lambda\rho} - \tfrac{1}{2}\epsilon_{\alpha\mu}\epsilon_{\beta\nu}\epsilon_{\lambda\rho}\gamma^{\alpha\beta}\gamma^{\lambda\rho} \\ &+ \tfrac{1}{8}\epsilon_{\mu\nu}\epsilon_{\alpha\beta}\epsilon_{\lambda\rho}\gamma^{\alpha\beta}\gamma^{\lambda\rho} - \tfrac{1}{4}\epsilon_{\mu\nu}\epsilon_{\alpha\lambda}\epsilon_{\beta\rho}\gamma^{\alpha\beta}\gamma^{\lambda\rho}) + 0(\kappa^3), \end{aligned} \tag{17}$$

where $\epsilon_{\mu\nu}$ is reciprocal to $\epsilon^{\mu\nu}$, and $0(\kappa^3)$ contains third and higher powers of κ.

If we substitute (15), (16), and (17) in (12), the Lagrangian density $\mathfrak{L}$ becomes an infinite series in powers of κ, each term of which is simply a function of $\gamma^{\mu\nu}$ and the Minkowskian metrical tensors $\epsilon^{\mu\nu}$ and $\epsilon_{\mu\nu}$. Thus, we obtain a flat-space expansion for $\mathfrak{L}$, which can be more conveniently expressed by using the usual flat space notation. In this way we find

$$\begin{aligned} \mathfrak{L} = -\frac{1}{4}&\left[\frac{\partial\gamma_{\mu\nu}}{\partial x_\lambda}\frac{\partial\gamma_{\mu\nu}}{\partial x_\lambda} - \frac{1}{2}\frac{\partial\gamma_{\alpha\alpha}}{\partial x_\lambda}\frac{\partial\gamma_{\beta\beta}}{\partial x_\lambda} - 2\frac{\partial\gamma_{\alpha\mu}}{\partial x_\nu}\frac{\partial\gamma_{\alpha\nu}}{\partial x_\mu}\right] \\ &+ \kappa\gamma_{\mu\nu}\left[\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu}\frac{\partial\gamma_{\alpha\beta}}{\partial x_\nu} - \frac{1}{2}\frac{\partial\gamma_{\alpha\alpha}}{\partial x_\mu}\frac{\partial\gamma_{\beta\beta}}{\partial x_\nu} + 2\frac{\partial\gamma_{\mu\alpha}}{\partial x_\beta}\frac{\partial\gamma_{\nu\beta}}{\partial x_\alpha}\right. \\ &\left. + \frac{\partial\gamma_{\beta\beta}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\alpha} - 2\frac{\partial\gamma_{\mu\beta}}{\partial x_\alpha}\frac{\partial\gamma_{\nu\beta}}{\partial x_\alpha}\right] + 0(\kappa^2). \end{aligned} \tag{18}$$

[3] G. D. Birkhoff, Proc. Natl. Acad. Sci. U. S. **29**, 231 (1943) and **30**, 324 (1944). See also M. Moshinsky, Phys. Rev. **80**, 514 (1950).
[4] H. Weyl, Am. J. Math. **66**, 591 (1944).
[5] S. N. Gupta, Phys. Rev. **96**, 1683 (1954).

[6] For mathematical details of the contents of this section, see S. N. Gupta, Proc. Phys. Soc. (London) **A65**, 608 (1952).

The above expansion can, of course, be obtained up to any desired power of κ. It is also possible to show that in flat space the gravitational field equation can be expressed as

$$\square^2\gamma_{\mu\nu}=\kappa(T_{\mu\nu}+t_{\mu\nu}) \tag{19}$$

with the supplementary condition

$$\partial\gamma_{\mu\nu}/\partial x_\nu=0, \tag{20}$$

where $T_{\mu\nu}$ and $t_{\mu\nu}$ are the symmetrical energy-momentum tensors for the matter field and the gravitational field respectively.[7] The extremely simple appearance of the field equation (19) is rather deceptive, because $t_{\mu\nu}$ in fact consists of an infinite series in powers of κ.

Comparing (19), (20), and (18) with (9), (6), and (11), we find that Einstein's field in flat space has just those properties, which a Lorentz-covariant gravitational field of spin 2 has to satisfy. Thus, we have shown that Einstein's theory can be reduced to a theory of gravitation in flat space by an expansion of the gravitational Lagrangian density as an infinite series in powers of the gravitational coupling constant κ. From a mathematical point of view, the theory remains unchanged when we pass over from the Riemannian space to the flat space. But, from a philosophic point of view, the above procedure implies a departure from Einstein's ideas in some respects.

[7] It is well known that in the Riemannian space the quantity $t_{\mu\nu}$ is not a tensor. However, when we pass over to the flat space and confine ourselves to the Lorentz transformations, we can regard $t_{\mu\nu}$ as a tensor.

IV. QUANTIZATION OF THE GRAVITATIONAL FIELD

We have seen in the preceding section that we can treat Einstein's theory as a theory of gravitation in flat space. Such a treatment has two great advantages. Firstly, it provides us with a more uniform description of the gravitational and the electromagnetic fields. Secondly, it enables us to carry out the quantization of Einstein's gravitational field by following the same procedure as we use for the electromagnetic field, as has been shown by the author.[8] On quantization, Einstein's gravitational field corresponds to gravitational quanta or gravitons of vanishing rest-mass and spin 2, and it is possible to calculate the interaction of these gravitons and other particles in the usual way. In such a quantized theory the nonlinearity of the gravitational field appears as a direct interaction between the gravitons.

An interesting application of the above theory of the quantization of the gravitational field has recently been carried out by Corinaldesi.[9] He has calculated the gravitational potential between two particles of spin 0 due to the exchange of a graviton, and he has then used this potential to find the two-body equation of motion under the influence of the mutual gravitational fields of these bodies. In this way he has derived in a remarkably straightforward way exactly the same equations of motion as obtained by Einstein, Infeld, and Hoffmann.[10]

Thus, the treatment of Einstein's gravitational field as a field in flat space and its quantization are also very useful from a practical point of view.

It is a pleasure to thank Professor John A. Wheeler for valuable discussions.

[8] S. N. Gupta, Proc. Phys. Soc. (London) **A65**, 161, 608 (1952).
[9] E. Corinaldesi, Nuovo cimento **1**, 1289 (1955) and **2**, 168 (1955); Proc. Phys. Soc. (London) **A69**, 189 (1956).
[10] Einstein, Infeld, and Hoffmann, Ann. Math. **39**, 66 (1938).

Electricity and General Relativity

CORNELIUS LANCZOS
Dublin Institute for Advanced Studies, Dublin, Eire

1. INTRODUCTION

IN ITS mathematical aspects, Einstein's theory of gravity is based on two fundamental assumptions: (1) the four-dimensional space-time manifold is Riemannian; (this implies existence of a symmetric metrical tensor g_{ik}); (2) the fundamental Lagrangian of the action principle is linear in the curvature components: $L=R_{ik}g^{ik}$. The resulting field equations $R_{ik}=0$ gave full account of the gravitational phenomena but failed to include the electric and quantum phenomena. They also failed to furnish new viewpoints for the structural problems of matter since they did not allow singularity-free solutions which could have been correlated to any of the elementary particles realized in nature.[1]

Later attempts at generalizing the gravitational theory subjected the fundamental postulates (1) and (2) to close scrutiny. Einstein, and many other scientific workers following his lead, was unwilling to depart from the linear Lagrangian of the gravitational equations, since any other choice of the Lagrangian leads to field equations of fourth order, in contradiction to the fact that the basic field equations of mathematical physics do not surpass the order two. A further difficulty was that departure from Einstein's Lagrangian seemed to put the equivalence principle in jeopardy. Hence, it seemed advisable to maintain Einstein's fundamental invariant but to abandon some of the limitations of Riemannian geometry. One could base the geometry of nature on the Γ quantities of an "affine connection" (Eddington, Einstein, Schroedinger), or on a combination of the $\Gamma_{ik}{}^{m}$ and the g_{ik}, without demanding their symmetry with respect to i, k; (Einstein).[2]

The author's own attempts are characterized by a different departure.[3] He endeavored to leave the Riemannian geometry unchanged and replaced Einstein's linear invariant by an invariant which is *quadratic* in the curvature components. A similar attempt was made earlier by H. Weyl, but in conjunction with a modification of Riemannian geometry in favor of a more general infinitesimal geometry.[4] The author showed that integration of the field equations obtained from the (purely Riemannian) quadratic action principle gives rise to a vectorial function which has the classical properties of the electromagnetic vector potential. However, the ensuing mathematical difficulties frustrated all advance beyond the linear approximation, and the theory remained in a rudimentary stage.

Recent advances in the general Hamiltonization of field equations led to a new mathematical method which is now fully adequate to the investigation of the quadratic action principle. It can now be demonstrated that Riemannian geometry, without any encroachments, contains the entire edifice of classical electromagnetism, together with the interrelation of gravitational and electromagnetic forces. The additional non-Maxwellian terms shed new light on the nature of elementary particles, conceived as static and singularity-free solutions of the fundamental field equations.

2. HAMILTONIZATION OF FIELD EQUATIONS

In every Lorentz-invariant variational field theory the fundamental conservation laws of momentum and energy appear in the form that the Minkowskian divergence of a certain symmetric tensor of second order, the "stress-energy tensor," vanishes, in consequence of the field equations.[5] We have to assume that the basic Lagrangian does not contain the field quantities in higher than *first* derivatives. In two fundamental cases; namely, Maxwell's equations and Dirac's equation of the electron, the basic Lagrangian is even *linear* in the first partial derivatives. The resultant field equations are then of not higher than first order, in analogy to the Hamiltonian equations of dynamics. In both cases the resulting system is linear in the derivatives, with constant coefficients. Since in the dynamical case any arbitrary Lagrangian can be transformed into the Hamiltonian canonical form, should it not be likewise possible to transform an arbitrary set of *partial* differential equations, deducible from a Lagrangian, into a normal form, corresponding to Hamilton's canonical form?

To answer this question, we first interpret Hamilton's procedure in a manner which is somewhat different from the traditional approach, based on Legendre's transformation.[6] The new formulation has the advantage of much greater flexibility and is directly applicable

[1] A. Einstein, Rev. Univ. nac. Tucuman **2**, 11 (1941).

[2] For a brief account of the very extensive literature and a condensed bibliography see Sir E. Whittaker, *History of the Theories of Aether and Electricity* (Nelson and Sons, London, 1953), pp. 188–192; see also, Cornelius Lanczos, Nuovo cimento Suppl. 2, 1193 (1955).

[3] Cornelius Lanczos, Phys. Rev. **39**, 716 (1932); **61**, 713 (1942); (subsequently quoted as I and II); Revs. Modern Phys. **21**, 497 (1949).

[4] Hermann Weyl, Math. Z. **2**, 384 (1918); Ann. Physik **59**, 101 (1919); Physik. Z. **22**, 473 (1921); see also W. Pauli, Enc. math. Wiss. V19, 759 (1920). In later years, under the impact of wave mechanics, Weyl lost confidence in his theory; see *Selecta Hermann Weyl* (Birkhäuser Verlag, Basel, 1956), p. 192.

[5] See, e.g., G. Wentzel, *Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1949), pp. 10 and 217.

[6] See C. Lanczos, *The Variational Principles of Mechanics* (Toronto University Press, Toronto, 1949), p. 161.

to the realm of partial differential equations, which is our present concern. The "momenta" p_i appear in this formulation as Lagrangian multipliers.

Considering the Lagrangian $L_0(q_i,\dot{q}_i,t)$ of a dynamical problem, we want to conceive the $\dot{q}_i$ as a second set of independent variables, let us say w_i. This is permissible, provided that in the variation we do not violate the condition

$$\dot{q}_i - w_i = 0. \quad (2.1)$$

Thus we have the Lagrangian $L_0(q_i,w_i,t)$ with the n auxiliary conditions (2.1). According to the usual procedure we multiply every auxiliary condition by an undetermined multiplier p_i and add it to the given Lagrangian. We thus get the modified Lagrangian

$$L = p_i(\dot{q}_i - w_i) + L_0(q_i,w_i,t). \quad (2.2)$$

The auxiliary conditions can be dropped since they appear as the consequence of the variational principle, varying with respect to the p_i, which are full-fledged new mechanical variables. The original Lagrangian problem is thus replaced by a new Lagrangian problem whose variables are p_i, q_i, w_i. The new Lagrangian has the "canonical form"

$$L = p_i\dot{q}_i - H, \quad (2.3)$$

where

$$H = p_i w_i - L_0(q_i,w_i,t). \quad (2.4)$$

This H is free of all derivatives. Moreover, the variables w_i appear solely in H. Hence they are purely *algebraic* variables which can be eliminated without any integration. The Euler equations applied to these variables give

$$p_i - (\partial H/\partial w_i) = 0. \quad (2.5)$$

Although these equations hold only for the actual motion, while their use for the elimination of the w_i means that we impose them also on the varied motion, this move is actually justified. The elimination involves no derivatives. Hence, the condition of "varying between definite limits" is not violated.

Eliminating the w_i from (2.5) we now obtain them as some explicit functions of p_j, q_j, t:

$$w_i = f_i(p_j,q_j,t). \quad (2.6)$$

We introduce these functions into (2.4) and arrive at a new form of H which is purely a function of the p_i, q_i, t:

$$H = H(p_i,q_i,t) \quad (2.7)$$

This construction of H is exactly the traditional one, except that the $\dot{q}_i$ have been replaced by the w_i which, however, is irrelevant since the final H does not depend on the w_i.

This method is applicable to ordinary as to partial differential equations. With the help of it any system of field equations derivable from a variational principle can be brought into a canonical form. Our present aim is restricted to the field equations of general relativity.

We thus start with a Lagrangian shaped to the demands of general relativity. We wish to see how the general canonical process operates under these circumstances. In Einstein's case the fundamental Lagrangian is $L_0 = R_{ik}g^{ik}$. However, we do not want to restrict ourselves to this special case since our aim is to study a Lagrangian which is formed of the curvature quantities in a *quadratic* rather than linear manner; (see Sec. 5). We thus leave the specific form of the Lagrangian free and assume that it is some function of the R_{ik} and the g^{ik}:

$$L_0 = f(R_{ik}, g^{ik}). \quad (2.8)$$

The tensor R_{ik} is a complicated differential operator of second order of the basic variables g_{ik}. We replace, however, R_{ik} by the algebraic variables w_{ik}, considering the equation

$$R_{ik} - w_{ik} = 0 \quad (2.9)$$

as an auxiliary condition of the variational problem. This gives the new Lagrangian

$$L = p^{ik}(R_{ik} - w_{ik}) - L_0(w_{ik}, g^{ik}). \quad (2.10)$$

Furthermore, we conceive R_{ik} as a differential operator of only *first* order by introducing the new field variables $\Gamma_{ik}{}^m$:

$$R_{ik} = \tfrac{1}{2}\left(\frac{\partial\Gamma_{i\alpha}{}^{\alpha}}{\partial x_k} + \frac{\partial\Gamma_{k\alpha}{}^{\alpha}}{\partial x_i}\right) - \frac{\partial\Gamma_{ik}{}^{\alpha}}{\partial x_\alpha} + \Gamma_{i\alpha}{}^{\beta}\Gamma_{k\beta}{}^{\alpha} - \Gamma_{ik}{}^{\beta}\Gamma_{\alpha\beta}{}^{\alpha}. \quad (2.11)$$

This is permissible, provided that we add as auxiliary conditions the equations which establish the $\Gamma_{ik}{}^m$ as functions of the g^{ik}:

$$\frac{\partial g^{ik}}{\partial x_m} + \Gamma_{\alpha m}{}^{i} g^{\alpha k} + \Gamma_{\alpha m}{}^{k} g^{\alpha i} = 0. \quad (2.12)$$

At this stage our Lagrangian appears in the following form:

$$L = \tfrac{1}{2}p^{ik}\left(\frac{\partial\Gamma_{i\alpha}{}^{\alpha}}{\partial x_k} + \frac{\partial\Gamma_{k\alpha}{}^{\alpha}}{\partial x_i} - 2\frac{\partial\Gamma_{ik}{}^{\alpha}}{\partial x_\alpha}\right) + \gamma_{ik}{}^{m}\frac{\partial g^{ik}}{\partial x_m} - H_1 - H_0 \quad (2.13)$$

where

$$H_1 = p^{ik}(\Gamma_{ik}{}^{\beta}\Gamma_{\beta\alpha}{}^{\alpha} - \Gamma_{i\alpha}{}^{\beta}\Gamma_{k\beta}{}^{\alpha}) - \gamma_{ik}{}^{m}(\Gamma_{\alpha m}{}^{i}g^{\alpha k} + \Gamma_{\alpha m}{}^{k}g^{\alpha i}) \quad (2.14)$$

$$H_0 = p^{ik}w_{ik} - f(w_{ik}, g^{ik}). \quad (2.15)$$

The field quantities of the variational problem are the g^{ik}, p^{ik}, $\Gamma_{ik}{}^m$, $\gamma_{ik}{}^m$ and the w_{ik}, all quantities which are symmetric in i, k. However, the w_{ik} can be eliminated with the help of the equations

$$p^{ik} = \partial f/\partial w_{ik}. \quad (2.16)$$

Substituting the result of this elimination in (2.15) we obtain H_0 as a function of g^{ik}, p^{ik}:

$$H_0=H_0(p^{ik},g^{ik}). \qquad (2.17)$$

Hence our field variables are: the 10 g^{ik} and 10 p^{ik}; the 40 $\Gamma_{ik}{}^m$ and 40 $\gamma_{ik}{}^m$, altogether 100 variables. In terms of an abstract "configuration space" we could conceive these variables as the components $\psi_1, \psi_2, \cdots \psi_N$ of a "vector" in a space of $N=100$ dimensions.

The canonical integrand (2.13) contains the partial derivatives of the field variables in a particularly simple, namely purely *linear* form. The highly nonlinear character of the relativistic equations is thrown completely into the Hamiltonian function $H=H_0+H_1$, but even here H_1 is of not higher than *third* degree in the field variables.

In terms of the abstract vector ψ, the canonical integrand may be written in the following homogeneous fashion:

$$L=\psi_i\alpha_{ik}{}^m\frac{\partial\psi_k}{\partial x_m}-H(\psi_1,\psi_2,\cdots\psi_N) \qquad (2.18)$$

where the $\alpha_{ik}{}^m$ indicate a set of numerical matrices. The lower indices i, k belong to the configuration space of the ψ_i, while the upper index m belongs to the space of the x_m. The term in which the $\alpha_{ik}{}^m$ appears is variationally equivalent to

$$\tfrac{1}{2}\alpha_{ik}{}^m\left(\psi_i\frac{\partial\psi_k}{\partial x_m}-\psi_k\frac{\partial\psi_i}{\partial x_m}\right). \qquad (2.19)$$

This shows that the matrices $\alpha_{ik}{}^m$ can be conceived as anti-symmetric in i, k. If the operator $\partial/\partial x_m$ is replaced by the self-adjoint operator $\partial/i\partial x_m$, the matrices $i\alpha_{ik}{}^m$ become purely imaginary. Addition of a real symmetric part would not change anything since we have added not more than a pure divergence which is variationally deletable. Thus generally we may conceive the matrices $i\alpha_{ik}{}^1, \cdots i\alpha_{ik}{}^4$ as 4 given numerical Hermitian matrices. In Dirac's case of the electron they become 4 by 4 matrices while in the case of general relativity we get a system of four 100 by 100 matrices; (although in a given particular problem great simplifications can take place, of course).

Even such a highly complicated system of nonlinear field equations as the field equations of general relativity may still be written in the form of a Dirac equation, if the four special Dirac matrices are replaced by a much more elaborate but still purely numerical set of matrices.[7]

[7] The analogy is still more pronounced on formulating the canonical system in terms of 50 *complex* field variables $p^{ik}+ig^{ik}$, and

$$\Gamma_{ik}{}^m+i(\gamma_{ik}{}^m-\tfrac{1}{3}\gamma_{i\alpha}{}^\alpha\delta_k{}^m-\tfrac{1}{3}\gamma_{k\alpha}{}^\alpha\delta_i{}^m).$$

The 100 real ψ_i components are thus reducible to 50 complex ψ_i, in harmony with the complex nature of wave-mechanical field variables.

3. FIELD EQUATIONS OF GENERAL RELATIVITY IN CANONICAL FORM

In view of the fact that the $\Gamma_{ik}{}^m$ quantities do not form a genuine tensor, one might think that the Lagrangian multipliers p^{ik} and $\gamma_{ik}{}^m$ will likewise lack tensor character. However, the determining equations for these field variables are deduced by *varying* the g^{ik} and the $\Gamma_{ik}{}^m$ and we know that *variationally* the $\Gamma_{ik}{}^m$ *do* behave like a genuine tensor, covariant in i, k, contravariant in m. For this reason the additional field variables p^{ik} and $\gamma_{ik}{}^m$ become genuine tensors of second, respectively third rank. Variation with respect to the $\Gamma_{ik}{}^m$ is particularly interesting since it does not involve the unknown function $H_0(p^{ik},g^{ik})$ which is free of the $\Gamma_{ik}{}^m$.

From now on only *covariant* operations appear in our deductions. For the purpose of denoting a covariant derivative we follow Einstein's procedure, although replacing his "semicolon" by a simple "comma," since a distinction between ordinary and covariant differentiation is not necessary. (In the few cases when ordinary derivatives are encountered, the usual notation $\partial/\partial x_i$ will be used.)

Variation of the $\Gamma_{ik}{}^m$ yields the following 40 equations:

$$p_{,m}{}^{ik}-\tfrac{1}{2}(p^{i\alpha}{}_{,\alpha}\delta_m{}^k+p^{k\alpha}{}_{,\alpha}\delta_m{}^i) +\gamma_{m\alpha}{}^i g^{\alpha k}+\gamma_{m\alpha}{}^k g^{\alpha i}=0. \qquad (3.1)$$

To this we add the 40 "conjugate" equations, obtained by varying the $\gamma_{ik}{}^m$:

$$\frac{\partial g^{ik}}{\partial x_m}+\Gamma_{m\alpha}{}^i g^{\alpha k}+\Gamma_{m\alpha}{}^k g^{\alpha i}=0. \qquad (3.2)$$

So far we have obtained 80 equations of the canonical system. The remaining 20 equations are obtained by varying with respect to the g^{ik} and the p^{ik}. The former variation yields

$$-\gamma_{ik}{}^\alpha{}_{,\alpha}=\frac{\partial H_0}{\partial g^{ik}}-\tfrac{1}{2}H_0 g_{ik}. \qquad (3.3)$$

The latter variation yields

$$\tfrac{1}{2}\left(\frac{\partial\Gamma_{i\alpha}{}^\alpha}{\partial x_k}+\frac{\partial\Gamma_{k\alpha}{}^\alpha}{\partial x_i}-2\frac{\partial\Gamma_{ik}{}^\alpha}{\partial x_\alpha}\right) +\Gamma_{i\alpha}{}^\beta\Gamma_{k\beta}{}^\alpha-\Gamma_{ik}{}^\beta\Gamma_{\beta\alpha}{}^\alpha=\frac{\partial H_0}{\partial p^{ik}}. \qquad (3.4)$$

All these equations are *linear* in the derivatives, with constant coefficients. The nonlinearity enters only *algebraically*, in the terms which do not contain derivatives, in conformity with the canonical equations of ordinary dynamics.

4. EINSTEIN'S LINEAR ACTION PRINCIPLE

We first encounter the canonical system in a 1925 paper of Einstein[8], when he became interested in the

[8] A. Einstein, Sitzber. preuss. Akad. Wiss. 414 (1925).

theory of "affine connection," as advocated by A. S. Eddington.[9] Einstein proposes[10] not to discard the metrical tensor in favor of the $\Gamma_{ik}{}^m$ quantities, but to determine both the g_{ik} and the $\Gamma_{ik}{}^m$ from a common variational principle in which the g_{ik} and $\Gamma_{ik}{}^m$ are treated as independent field variables, but not assuming their symmetry with respect to i, k. He remarks that for the case of pure gravity (i.e., symmetric g_{ik} and $\Gamma_{ik}{}^m$), one obtains the most satisfactory derivation of the ordinary gravitational equations $R_{ik}=0$. Einstein does not use any Lagrangian multipliers and his system (for the case of symmetry) contains only 50 instead of 100 field variables. Why is it that in his case the additional variables do not come into evidence?

Einstein chose the action principle

$$L_0=R_{ik}g^{ik}. \tag{4.1}$$

In this case

$$f(w_{ik},g^{ik})=w_{ik}g^{ik} \tag{4.2}$$

and

$$H_0=(p^{ik}-g^{ik})w_{ik}. \tag{4.3}$$

In view of the linearity of H_0 in w_{ik} we do not succeed now with the elimination scheme. Nor is that necessary in the present case. Addition of H_0 to the Lagrangian can be conceived as a tool to maintain the auxiliary condition

$$p^{ik}=g^{ik}. \tag{4.4}$$

By introducing this condition in (2.13) we dispense with the field variables p^{ik} and also with H_0, although the Lagrangian factors $\gamma_{ik}{}^m$ still remain. But now (3.1) yields for the present case a complete *vanishing* of the $\gamma_{ik}{}^m$, since the covariant derivatives of the g_{ik} are zero. By maintaining the condition $\gamma_{ik}{}^m=0$ during the variation we lose the Lagrangian multipliers and (3.2), but we are entitled to do so because (3.1) take over the role of the equations (3.2).

From the standpoint of the general theory the case of Einstein's linear action principle represents a *degenerate* case in which the 50 conjugate field variables become tautological. The objection that any other choice of the action principle leads to differential equations of fourth order for the g_{ik}, obtains a different meaning on viewing the problem from the vantage point of the canonical equations. The canonical equations remain of first order under all circumstances, and it is only the vanishing of the conjugate variables which characterizes the Einsteinian system. Since, however, the 100 field variables of the general scheme are replaceable by 50 complex field variables,[7] we can regard the action principle of Einstein as that extreme case in which the generally complex field variables are reduced to real variables.

[9] A. S. Eddington, Proc. Roy. Soc. (London) **A99**, 104 (1921).

[10] The general program of this paper is remarkably close to the last efforts of Einstein toward a unified field theory; see Albert Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, 1955), fifth edition, p. 154.

5. QUADRATIC ACTION PRINCIPLE

We now depart from Einstein's linear action principle and introduce the following Lagrangian[11]

$$\begin{aligned} L_0&=\tfrac{1}{2}(R_{ik}R^{ik}+\beta R^2)\\ &=\tfrac{1}{2}(w_{ik}w^{ik}+\beta w^2), \end{aligned} \tag{5.1}$$

where β is an unspecified numerical constant. We have to eliminate w_{ik} from the Hamiltonian

$$H_0=p^{ik}w_{ik}-\tfrac{1}{2}(w^{ik}w_{ik}+\beta w^2). \tag{5.2}$$

The elimination yields

$$H_0=\tfrac{1}{2}(p^{ik}p_{ik}+\sigma p^2) \tag{5.3}$$

with

$$\sigma=-\frac{\beta}{1+4\beta}. \tag{5.4}$$

We do not use the full canonical scheme of Sec. 2, in order to facilitate comparison with the classical results of general relativity. We allow differential operators of second order by conceiving R_{ik} as a direct function of the g_{ik}, without interjecting the $\Gamma_{ik}{}^m$ as an additional set of independent variables. Then also the conjugate $\gamma_{ik}{}^m$ will not appear. Our action variables are solely the 10 g_{ik} and the 10 p^{ik}. The Lagrangian (2.13) will therefore simplify to

$$\begin{aligned} L&=p^{ik}R_{ik}-H_0\\ &=p^{ik}R_{ik}-\tfrac{1}{2}(p^{ik}p_{ik}+\sigma p^2), \end{aligned} \tag{5.5}$$

where R_{ik} is defined by (2.11) on substituting for the $\Gamma_{ik}{}^m$ the usual expressions in terms of the g_{ik} and their first derivatives. Hence from now on R_{ik} will be regarded as a second-order differential operator of the g_{ik}.

In order to perform the variation with respect to the g_{ik} and the p_{ik}, we need a certain invariant differential operator of second order, generated by variation of R_{ik}. This expression has been derived previously[12] and may be written as follows:

$$E_{ik}(\gamma)=\delta R_{ik}=\tfrac{1}{2}(\Delta\gamma_{ik}-\gamma_i{}^\alpha{}_{,k\alpha}-\gamma_k{}^\alpha{}_{,i\alpha}+\gamma_\alpha{}^\alpha{}_{,ik}) \tag{5.6}$$

where $\gamma_{ik}=\delta g_{ik}$ and Δ denotes the invariant Laplace operator

$$\Delta u_{i\cdots p}=g^{\alpha\beta}u_{i\cdots p,\,\alpha\beta}. \tag{5.7}$$

We also need the adjoint of this operator, likewise given before [see I (2.10)]:

$$D_{ik}(p)=\tfrac{1}{2}(\Delta p_{ik}-p_i{}^\alpha{}_{,k\alpha}-p_k{}^\alpha{}_{,i\alpha}+p^{\alpha\beta}{}_{,\alpha\beta}g_{ik}). \tag{5.8}$$

In terms of this operator the variation with respect to g_{ik} yields:

$$D^{ik}(p)-(p^{i\alpha}p_\alpha{}^k+\sigma pp^{ik})+\tfrac{1}{2}Lg^{ik}=0 \tag{5.9}$$

while the variation with respect to p^{ik} gives

$$R_{ik}=p_{ik}+\sigma pg_{ik}. \tag{5.10}$$

[11] Concerning the uniqueness of this choice see II, p. 714.

[12] See I (2.6) and (2.12). The present E corresponds to $\tfrac{1}{2}D$, the present D to $\tfrac{1}{2}F$ of the previous paper.

In consequence of (5.10) we may put

$$L=H_0. \tag{5.11}$$

On the basis of this relation (5.9) may be written as follows:

$$D^{ik}(p)-(p^{i\alpha}p_\alpha{}^k-\tfrac{1}{4}p^{\alpha\beta}p_{\alpha\beta}g_{ik})-\sigma p(p^{ik}-\tfrac{1}{4}pg^{ik})=0. \tag{5.12}$$

The 20 equations (5.10) and (5.12) are the fundamental field equations of our problem.

With the help of these equations we demonstrate two fundamental consequences of the quadratic action principle, obtained earlier by different tools; [see II (2.13) and (2.15)]. Replace the notation p_{ik} by $\bar{p}_{ik}$ and put

$$\bar{p}_{ik}=p_{ik}+\lambda g_{ik}. \tag{5.13}$$

Then Eq. (5.10) changes to

$$R_{ik}=(1+4\sigma)\lambda g_{ik}+p_{ik}+\sigma p g_{ik} \tag{5.14}$$

while the modification of the left side of (5.12) is only the addition of the following term:

$$-2\lambda(1+2\sigma)(p^{ik}-\tfrac{1}{4}pg^{ik}). \tag{5.15}$$

The inhomogeneous part of the substitution drops out completely of the (5.12) and $p_{ik}=0$ is still a possible solution of the field equations. This means that the so-called "cosmological equations"

$$R_{ik}=(1+4\sigma)\lambda g_{ik} \tag{5.16}$$

with an arbitrary constant λ represent an *exact solution* of the field equations. Any solution of (5.12), [corrected by the term (5.15)], which is not zero but small, can be conceived as a small deformation of the fundamental cosmological solution (5.16). The constant λ, usually considered as completely negligible and modifying only the structure of the world in cosmic dimensions, will in our later discussions play the role of a fundamental *atomic* constant; (see Sec. 9).

As a second consequence, multiply (5.12) by g_{ik}. We get

$$\Delta p+2p^{\alpha\beta}{}_{,\alpha\beta}=0. \tag{5.17}$$

On the other hand, taking the divergence of the divergence condition for the "metrical matter tensor"

$$T_{ik}=R_{ik}-\tfrac{1}{2}Rg_{ik} \tag{5.18}$$

we obtain the relation

$$R^{\alpha\beta}{}_{,\alpha\beta}=\tfrac{1}{2}\Delta R \tag{5.19}$$

which means, in view of (5.10)

$$p^{\alpha\beta}{}_{,\alpha\beta}=\tfrac{1}{2}(1+2\sigma)\Delta p. \tag{5.20}$$

Substitution in (5.17) gives

$$(1+\sigma)\Delta p=0. \tag{5.21}$$

which may also be written in the form

$$\frac{1+\sigma}{1+4\sigma}\Delta R=0. \tag{5.22}$$

We see that $R=\text{const}$ is an *exact first integral* of the field equations. The special choice $\sigma=-1$ of the free constant σ leads to a *degeneracy*. We lose one of the determining equations of our system and thus the solution becomes under-determined; (cf. Sec. 7).

6. FREE VECTORIAL FUNCTION φ_i

In problems of Hamiltonian dynamics the theory of canonical transformations is frequently of paramount importance.[13] Instead of trying to integrate the dynamical equations directly, we solve them indirectly by performing transformations of the dynamical variables which retain the normal form of the dynamical equations but simplify the Hamiltonian function H. The canonical nature of the transformation demands that the differential form $p_i\dot{q}_i$ shall change by a complete derivative only. In this case the canonical equations are preserved but the transformation modifies the form of the Hamiltonian function H.

In our present problem the derivative part of the Lagrangian (5.5) is contained in the first term. A canonical transformation of the variables p_{ik} and g_{ik} should have the property that it should change the term $p^{ik}R_{ik}$ by a pure divergence. The new Lagrangian will then be of the same form as the original one, but with a modified Hamiltonian H. We will now take advantage of the fact that the divergence of the metrical matter tensor (5.18) *vanishes identically*. We denote once more the original p_{ik} by $\bar{p}_{ik}$ and apply the following transformation:

$$\bar{p}_{ik}=p_{ik}+\tfrac{1}{2}(\varphi_{i,k}+\varphi_{k,i}-\varphi^\alpha{}_{,\alpha}g_{ik}). \tag{6.1}$$

A simple calculation shows that the added terms contribute to L a pure divergence which can be omitted. Hence our new Lagrangian becomes

$$L=p^{ik}R_{ik}-H, \tag{6.2}$$

where H, compared with the previous H_0, is augmented by further terms, partly linear and partly quadratic in φ_i. First consider the *quadratic* terms only. They amount to the following addition H_1 to the previous H_0[14]:

$$H_1=\tfrac{1}{8}(\varphi_{i,k}+\varphi_{k,i})(\varphi^{i,k}+\varphi^{k,i})+\tfrac{1}{2}\sigma(\varphi^\alpha{}_{,\alpha})^2. \tag{6.3}$$

The first term can be transformed as follows:

$$\begin{aligned}\tfrac{1}{8}(\varphi_{i,k}+\varphi_{k,i})(\varphi^{i,k}+\varphi^{k,i})&=\tfrac{1}{4}(\varphi_{i,k}+\varphi_{k,i})\varphi^{i,k}\\&=\tfrac{1}{4}(\varphi_{i,k}-\varphi_{k,i})\varphi^{i,k}+\tfrac{1}{2}\varphi_{k,i}\varphi^{i,k}\\&=\tfrac{1}{8}F_{ik}F^{ik}+\tfrac{1}{2}\varphi_{k,i}\varphi^{i,k},\end{aligned} \tag{6.4}$$

[13] See reference 6, Chap. VII.
[14] The notation $\varphi^{i,k}$ refers to $\varphi^i{}_{,\alpha}g^{\alpha k}$.

where

$$F_{ik}=\varphi_{i,k}-\varphi_{k,i}=\frac{\partial\varphi_i}{\partial x_k}-\frac{\partial\varphi_k}{\partial x_i}. \quad (6.5)$$

Moreover,

$$\begin{aligned}\varphi^{i,k}\varphi_{k,i}&=(\varphi^{i,k}\varphi_k)_{,i}-\varphi^{i,k}{}_{,i}\varphi_k\\&=(\varphi^{i,k}\varphi_k)_{,i}-\varphi^{i}{}_{,ik}\varphi^k+R_{ik}\varphi^i\varphi^k\\&=(\varphi^{i}{}_{,k}\varphi^k-\varphi^{k}{}_{,k}\varphi^i)_{,i}+(\varphi^{\alpha}{}_{,\alpha})^2+R_{ik}\varphi^i\varphi^k. \quad (6.6)\end{aligned}$$

Omitting a divergence, which is variationally zero, we can write

$$H_1=\tfrac{1}{8}F_{ik}F^{ik}+\tfrac{1}{2}(1+\sigma)(\varphi^{\alpha}{}_{,\alpha})^2+\tfrac{1}{2}R_{ik}\varphi^i\varphi^k. \quad (6.7)$$

The last term contains once more the differential operator R_{ik} which should be united with the first term. Hence, our final transformation will not be (6.1) but

$$\bar{p}_{ik}=p_{ik}+\tfrac{1}{2}(\varphi_{i,k}+\varphi_{k,i}-\varphi^{\alpha}{}_{,\alpha}g_{ik}+\varphi_i\varphi_k). \quad (6.8)$$

We know in advance that the term $p^{ik}R_{ik}$ will remain unchanged. The Hamiltonian H_0, however, has to be augmented by further terms which will be partly linear in the p^{ik} and partly independent of them. Those terms which are linear in the p^{ik}, can be combined with the first term of the Lagrangian by introducing a modified differential operator R_{ik}^*. Then the Hamiltonian H_1 has to absorb only those terms which depend on φ_i alone. Carrying through the calculations, we arrive at the following result:

$$L=p^{ik}R_{ik}^*-(H_0+H_1), \quad (6.9)$$

where

$$R_{ik}^*=R_{ik}-\tfrac{1}{2}(\varphi_{i,k}+\varphi_{k,i}+\varphi_i\varphi_k)+\tfrac{1}{2}[(1+2\sigma)\varphi^{\alpha}{}_{,\alpha}-\sigma\varphi^{\alpha}\varphi_{\alpha}]g_{ik} \quad (6.10)$$

and

$$H_1=\tfrac{1}{8}F_{ik}F^{ik}+\tfrac{1}{2}(1+\sigma)(\varphi^{\alpha}{}_{,\alpha}-\tfrac{1}{2}\varphi^{\alpha}\varphi_{\alpha})^2 \quad (6.11)$$

while H_0 is still defined by (5.3).

The transformation (6.8) correlates to any $\bar{p}_{ik}$ a new p_{ik}, considering φ_i as an arbitrarily prescribed (although continuous and differentiable) vector function. Varying with respect to the p_{ik} and the g_{ik}, considering the φ_i as given, we obtain the previous field equations, although now expressed in the new p_{ik}. What happens if we include the φ_i among the field variables by adding their free variation to the free variations of p_{ik} and g_{ik}? The transformation equation (6.8) shows that if we keep p_{ik} constant but vary φ_i arbitrarily, we obtain a certain special type of variation for the $\bar{p}_{ik}$. The field equations guarantee the vanishing of the first variation of our action integral for *any* variation of the $\bar{p}_{ik}$. Hence addition of the φ_i to our field variables merely adds four more equations which are tautological since they are satisfied in consequence of the field equations. Thus we do not lose in generality if we add the φ_i to the field variables of our problem. The 24 field equations thus derived do not yield in fact more than 20 independent equations, obtained by varying the p_{ik} and the g_{ik}.

7. THEORY OF WEYL

We now discuss a remarkable relation between our purely Riemannian action principle and the theory of Hermann Weyl,[4] developed on the basis of an infinitesimal geometry which is more general than Riemann's geometry. Weyl constructed a geometry which enlarges the Riemannian line-element g_{ik} by a further vectorial quantity φ_i, which in physical interpretation becomes the electromagnetic vector potential. The fundamental field equations of this geometry are developed from an action principle which is quadratic in the curvature quantities, in full analogy to the assumptions of the present theory. The general quadratic action principle was considered by W. Pauli,[15] who developed all the relevant mathematical and physical consequences of Weyl's theory. It is composed of a linear superposition of four terms, with three arbitrary constants, and may be written, according to Pauli, in the following form [see Pauli's Eq. (32), p. 462]:

$$L_p=-\tfrac{1}{2}\Big[k_1\tilde{R}_{iklm}\tilde{R}^{iklm}+k_2\tilde{R}_{ik}\tilde{R}^{ik}+\frac{k_3}{2}\tilde{R}^2+\tfrac{1}{4}F_{ik}F^{ik}\Big], \quad (7.1)$$

where

$$\tilde{R}_{ik}=R_{ik}-\tfrac{1}{2}(\varphi_{i,k}+\varphi_{k,i}+\varphi_i\varphi_k)-\tfrac{1}{2}(\varphi^{\alpha}{}_{,\alpha}-\varphi^{\alpha}\varphi_{\alpha})g_{ik}. \quad (7.2)$$

In fact it is unnecessary to include the term with $\tilde{R}_{iklm}$ since one can show[16] that variation of the first invariant is equivalent to that of a certain linear combination of the other three invariants. Hence, it is permissible to put in advance $k_1=0$, retaining only the two essential constants k_2 and k_3. In the Riemannian case, where a similar reduction takes place,[17] only *one* essential constant remains, *viz.* the β of (3.1).

To faciliate a comparison between our purely Riemannian Lagrangian (6.9) and Weyl's Lagrangian (7.1), we now eliminate the p_{ik} from our action principle, thus returning once more to the Lagrangian instead of Hamiltonian formulation of action. The elimination gives the new Lagrangian

$$L_0'=\tfrac{1}{2}(R_{ik}^*R^{*ik}+\beta R^{*2})-H_1 \quad (7.3)$$

in which only the g_{ik} and the φ_i remain as field variables. Comparison of the R_{ik}^* [see (6.10)] with the $\tilde{R}_{ik}$ of Weyl's theory shows a remarkable analogy. The general structure of the terms is exactly the same in both cases, except that the numerical coefficients within the last term do not agree. For the special choice $\sigma=-1$ the two expressions become *exactly* identical. But even for other values of σ the difference is of *second* order only, if we assume the customary "Lorentz

[15] W. Pauli, Physik. Z. 20, 457 (1919).
[16] See R. Bach, Math. Z. 9, 110 (1921).
[17] See C. Lanczos, Ann. Math. 39, 842 (1938).

condition" for the vector potential:

$$\varphi^{\alpha}{}_{,\alpha}=0. \qquad (7.4)$$

And yet, the origins of R_{ik}^* and $\bar{R}_{ik}$ are widely different. The first expression came about as the result of a canonical transformation (with φ_i=arbitrary), the second on the basis of a certain type of non-Riemannian geometry. Discarding for the moment the small difference between the two kinds of curvature quantities, Weyl's modification of Riemannian geometry appears tautological, except for the fact that Weyl's action principle permits *two* free constants k_2 and k_3, while the Riemannian case leads to the *single* free constant β. This greater freedom of Weyl's geometry is not necessarily an advantage since the two free constants k_2 and k_3 cannot be correlated to definite physical constants. Pauli arrives at a decision on the basis that the results of Einstein's gravitational theory should be maintained for the case of vanishing φ_i. He thus makes the choice $k_1=k_2=0$; [see his equation (50), p. 465]; Weyl subsequently adopted this choice.[18]

In the Riemannian case we do not have the freedom of k_2. The comparison of (7.3) with (7.1) shows that here

$$k_2=-1 \qquad (7.5)$$

while the constants k_3 and β are in the relation

$$k_3=-2\beta=2\sigma/(1+4\sigma). \qquad (7.6)$$

Although (7.5) prohibits the choice $k_2=0$, we can interpret Pauli's choice as the limiting case

$$k_2/k_3=0 \qquad (7.7)$$

which means

$$\beta=\infty, \quad \sigma=-\tfrac{1}{4}. \qquad (7.8)$$

There exists a deep-seated difference between the theory of Weyl and the theory here advocated. In Weyl's geometry the "principle of gauge-invariance" invokes a certain pre-established harmony between the g_{ik} and the φ_i, chosen in such manner that at every point of the manifold a proportionality factor of the g_{ik} must remain undetermined; correspondingly, the φ_i can also be determined up to a scalar function only. Einstein[19] found Weyl's theory unacceptable on account of the arbitrariness of a common factor of the g_{ik}. He pointed out that the tremendous consistency of the spectral lines throughout the universe demonstrates the inevitability of an absolute measuring rod. The present theory is not affected by this criticism of Einstein since the g_{ik} are in our case completely determined. On the other hand, the vector potential φ_i remains completely undetermined. This shows that the limiting case $\sigma=-1$, i.e.,

$$\beta=-\tfrac{1}{3}; \quad k_1=0, \quad k_2=-1, \quad k_3=\tfrac{2}{3} \qquad (7.9)$$

which is a point of tangency between the two theories, must be singular from both the Riemannian and the Weylian approach. From the Riemannian standpoint we must lose one of our equations in determining the metrical tensor, from the Weylian standpoint we must lose the determining equation for the vector potential.

Closer examination corroborates these predictions. We have seen in Sec. 5 that $\sigma=-1$ is a singular case for which the determining equation (5.22) of the scalar curvature R is lost. This is in harmony with Weyl's geometry in which R remains undetermined on account of the freedom of an arbitrary factor of the g_{ik}. Let us see, what we get for the vector potential in Weyl's theory. Variation of the vector potential in Pauli's action principle (7.1) yields the following relation[20]:

$$\tfrac{1}{2}(1+2k_1+k_2)\frac{\partial g^{\frac{1}{2}}F^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha} = (k_1+k_2+\tfrac{3}{2}k_3)(\bar{R}_{,\alpha}g^{\alpha i}+\bar{R}\varphi^i). \qquad (7.10)$$

For the case $k_1=0$, $k_2=-1$, the left side of the equation disappears. The right side would then demand the complete vanishing of φ_i, except if the choice $k_3=\frac{2}{3}$ is made, which makes the right side vanish too. This choice corresponds exactly to our singular case $\sigma=-1$, as (7.6) demonstrates. For other values of σ, the left side of (7.10) is still zero but the right side is now corrected by further terms because of the modifications which have to be applied to the quantities of Weyl. The exact relation between R_{ik}^* and $\bar{R}_{ik}$ is given as follows:

$$R_{ik}^*=\bar{R}_{ik}+(1+\sigma)Qg_{ik}, \qquad (7.11)$$

where the scalar Q is a certain non-Maxwellian quantity which plays a characteristic and fundamental role in the present theory

$$Q=\varphi^{\alpha}{}_{,\alpha}-\tfrac{1}{2}\varphi^\alpha\varphi_\alpha. \qquad (7.12)$$

Moreover, our Lagrangian (7.3) and Pauli's Lagrangian (7.1) (for the case $k_1=0$, $k_2=-1$, $k_3=-2\beta$) are in the following relation to each other:

$$L_0'=L_p+(1+3\beta)Q(\bar{R}+\tfrac{3}{2}\phi). \qquad (7.13)$$

The correction term on the right side is such that its variation with respect to φ_i exactly counteracts the right side of (7.10) and once more we obtain the empty equation $0=0$ which leaves φ_i undetermined.

8. VECTOR POTENTIAL AND METRICAL TENSOR

The vectorial function φ_i—interpreted in Weyl's theory as the vector potential of the electromagnetic field—appears in our investigation as a quantity which

[18] See H. Weyl, *Space, Time, Matter* (Methuen and Company, London, 1922), p. 295.
[19] Albert Einstein, Physik. Z. 21, 651 (1920).

[20] See Eqs. (26), (29), (30), and (32) of Pauli's paper; Pauli's equation (26) should on the right side contain the factor $(1+2c_1+c_2)$. The error is caused by the identity (*B*) on p. 460, which in fact holds for $\bar{R}^{ki}$ and not for $\bar{R}^{ik}$. At the author's request Professor Pauli kindly rechecked his calculations and corroborated the author's findings.

originates in a certain canonical transformation. The inner meaning of φ_i does not reveal itself by this method. We now approach the same problem differently to obtain further clues concerning the true significance of the vector potential.

One of the earliest investigations of Einstein[21] brought already to light a peculiar feature of the contracted curvature tensor. Integration of the gravitational equations succeeds in a natural way only if one normalizes the reference system in a very special manner, in spite of the general covariance of the fundamental equations. Einstein's normalization for infinitesimal fields was afterwards extended by the author[22] to fields of arbitrary strength. The normalization condition takes the form

$$\frac{1}{g^{\frac{1}{2}}}\frac{\partial g^{\frac{1}{2}}g^{i\alpha}}{\partial x_\alpha}=0. \tag{8.1}$$

The effect of this normalization is that the differential operator R_{ik} becomes reducible to a greatly simplified operator B_{ik} which has the merit that its essential part (which contains the second derivatives) is *separated* in the components g_{ik}. Moreover, this operator is free of any inner identities. It can be prescribed freely, without having to satisfy any conservation laws,—in marked contrast to the original operator R_{ik}.

The general expression of B_{ik} can be written down as follows:

$$B_{ik}=\tfrac{1}{2}\frac{\partial^2 g_{ik}}{\partial x_\alpha \partial x_\beta}g^{\alpha\beta}+\tfrac{1}{2}\Gamma_{\alpha\beta}{}^{\rho}\left(\frac{\partial g^{\alpha\beta}}{\partial x_k}g_{\rho i}+\frac{\partial g^{\alpha\beta}}{\partial x_i}g_{\rho k}\right) -\Gamma_{i\alpha}{}^{\rho}\Gamma_{k\beta}{}^{\sigma}g_{\rho\sigma}g^{\alpha\beta}. \tag{8.2}$$

Without restricting our reference system by any special condition we can say quite generally that the operator R_{ik} may be split in a definite noncovariant manner into the simplified operator B_{ik} plus the symmetrized gradient of a certain vectorial quantity. We can put

$$R_{ik}=B_{ik}+\tfrac{1}{2}(V_{i,k}+V_{k,i}) \tag{8.3}$$

where

$$V^i=\frac{1}{g^{\frac{1}{2}}}\frac{\partial g^{\frac{1}{2}}g^{i\alpha}}{\partial x_\alpha}. \tag{8.4}$$

(The operation "comma" is taken in the usual invariant sense, although $V_i=V^\alpha g_{\alpha i}$ is not a true vector.)

We introduce this B_{ik} in our discussions, replacing R_{ik} by the expression (8.3). We substitute (8.3) in our quadratic Lagrangian (5.1) and take into account the divergence-free character of the matter tensor (5.18). By a similar procedure as that employed in Sec. 6 we transform L_0 as follows:

$$L_0=\tfrac{1}{2}[B^{ik}B_{ik}+\beta B^2+(1+2\beta)BV-V^iV^kB_{ik} -\tfrac{1}{4}W^{ik}W_{ik}+\beta V^2+\tfrac{1}{2}VV^\alpha V_\alpha], \tag{8.5}$$

where we have put

$$V=V^\alpha{}_{,\alpha} \tag{8.6}$$

$$W_{ik}=\frac{\partial V_i}{\partial x_k}-\frac{\partial V_k}{\partial x_i}. \tag{8.7}$$

We now proceed to the canonical method by replacing B_{ik} and V^i by the algebraic variables w_{ik} and φ^i. This gives rise to the Lagrangian multipliers p^{ik} and ρ_i and we obtain the new Lagrangian

$$L=p^{ik}(B_{ik}-w_{ik})+\rho_i(V^i-\varphi^i)+L_0, \tag{8.8}$$

where L_0 is the above expression (8.5), but replacing B_{ik} by $w_i{}^\alpha$ and V_i by φ_i. We recognize that L is purely algebraic in the variables w_{ik} which can thus be eliminated. (The same is not true for φ_i which enters in L_0 with its first derivatives.) By putting the partial derivative of L with respect to w_{ik} equal to zero we obtain the equation

$$p^{ik}=w^{ik}+\beta w g^{ik}+\tfrac{1}{2}(1+2\beta)\varphi g^{ik}-\tfrac{1}{2}\varphi^i\varphi^k \tag{8.9}$$

from which w_{ik} can be eliminated. After the elimination the final Lagrangian becomes

$$L=p^{ik}B_{ik}^*-(H_0+H_1)+\rho_i\left(\frac{\partial g^{\frac{1}{2}}g^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}-\psi^i\right) \tag{8.10}$$

where

$$B_{ik}^*=B_{ik}-\tfrac{1}{2}\varphi_i\varphi_k+\tfrac{1}{2}[(1+2\sigma)\varphi^\alpha{}_{,\alpha}-\sigma\varphi^\alpha\varphi_\alpha]g_{ik} \tag{8.11}$$

while H_0 and H_1 are again defined by (5.3) and (6.11).

On comparing the new Lagrangian (8.10) with the expression (6.9) found in Sec. 6, we see the close resemblance. The last term of (8.10) was not present in the earlier treatment. Moreover, the definition of R_{ik}^* according to (6.10) contains, if compared with (8.11), the additional term $-\tfrac{1}{2}(\varphi_{i,k}+\varphi_{k,i})$, but then this term is absorbed by B_{ik} as (8.3) shows.

Variation with respect to φ_i gives an equation which can be interpreted as the Maxwellian equation for the vector potential, augmented by certain correction terms:

$$\tfrac{1}{2}\frac{\partial g^{\frac{1}{2}}F^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}+[(1+\sigma)Q-\tfrac{1}{2}(1+2\sigma)p]_{,\alpha}g^{\alpha i}+(1+\sigma)Q\varphi^i -(p^{i\alpha}+\sigma p g^{i\alpha})\varphi_\alpha=\rho^i. \tag{8.12}$$

How does the variational principle determine the right side of this equation, the "electric current"? Variation with respect to ρ_i gives the condition

$$\frac{\partial g^{\frac{1}{2}}g^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}=\varphi^i \tag{8.13}$$

which says that the normalization vector, usually put equal to zero, [see 8.1)], should in actual fact be equated to the vector potential. But what can we say about ρ_i which enters the Lagrangian in a linear way

[21] Albert Einstein, Sitzber. preuss. Akad. Wiss. 688 (1916).
[22] Cornelius Lanczos, Physik. Z. 23, 537 (1922); Z. Physik 13, 7 (1923).

only? It so happens that ρ_i is determined in an *indirect* way.

Let us express in the Lagrangian (8.10) the operator B_{ik} in terms of R_{ik} with the help of (8.3), then perform an integration by parts. We arrive exactly at the invariant (6.9), augmented by a term which has the form of the last term of (8.10), except that ρ_i is replaced by a ρ_i' which is related to ρ_i as follows:

$$\rho_i'=\rho_i+p_{i\ \ ,\alpha}^{\alpha}. \tag{8.14}$$

This Lagrangian satisfies the condition of general covariance, except for the single term

$$\rho_i'\frac{\partial g^{\frac{1}{2}}g^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}. \tag{8.15}$$

Let us vary with respect to g^{ik}. The variation of the completely covariant part of L yields a certain symmetric tensor A_{ik}. Adding the variation of the term (8.15), the resulting equation becomes

$$A_{ik}-\tfrac{1}{2}\left(\frac{\partial\rho_i'}{\partial x_k}+\frac{\partial\rho_k'}{\partial x_i}-\frac{\partial\rho_\alpha'}{\partial x_\beta}g^{\alpha\beta}g_{ik}\right)=0. \tag{8.16}$$

We take the covariant divergence of this equation. The divergence of A_{ik} vanishes identically from the general principles of tensor calculus. We thus obtain for ρ_i' a linear and homogeneous partial differential equation of second order. We can assume that this equation has no solution which is regular throughout the space-time manifold and vanishes at infinity. Hence we can put $\rho_i=0$ and we obtain

$$\rho^i=-p^{i\alpha}_{\ \ ,\alpha}. \tag{8.17}$$

The current vector ρ^i is thus uniquely determined.

We return to our Lagrangian (8.10) and vary with respect to p^{ik}. This gives the equation

$$B_{ik}^*=p_{ik}+\sigma p g_{ik}. \tag{8.18}$$

Then we vary with respect to the g_{ik}. We obtain a linear differential operator of second order in the p^{ik} which becomes equal to quantities *quadratic* in the φ_i. Integration of this equation establishes the p^{ik} in terms of the φ_i. The quadratic dependence has the consequence that for weak φ_i fields the p^{ik} become infinitesimal of second order. This leads to the conclusion that the non-Maxwellian terms of (8.12) drop to quantities of second and third order. Hence, for weak fields the determining equation of the vector potential becomes

$$\frac{\partial g^{\frac{1}{2}}F^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}=0, \tag{8.19}$$

in harmony with the Maxwellian equations of empty space. If in (8.18) we go to the same degree of approximation, we obtain for weak fields the simplified equation

$$B_{ik}=0. \tag{8.20}$$

Let us consider an electrostatic field. Here the vector potential is reduced to the scalar potential which decreases to zero with r^{-1}. Integration of the field equations offers series difficulties. Integration of the fourth equation of (8.13) would demand that the g_{i4} ($i=1,2,3$) go to infinity with the order x_i/r, in contradiction to the boundary condition that they have to vanish for $r=\infty$. Moreover, the corresponding B_{i4} quantities do not become zero and thus (8.20) are not satisfied. On the other hand, if we consider the centrally symmetric and static solutions of (8.20), we find that φ_4 vanishes identically. Existence of a free electric charge is thus in contradiction to the demands of infinitesimal fields. This difficulty was recognized by the author at a very early stage of his speculations (compare I, page 735). It was baffling that on the one hand the quadratic action principle seemed to offer the vector potential as a free gift, on the other it took it away by denying the possibility of a free charge. The proper explanation of the puzzle could not be found as long as the adequate mathematical method for the treatment of strong fields was missing. Discovery of the canonical method puts us in the position to see the wider implications of the problem. The proper role of the vector potential φ_i in relation to the structural problems of matter can now be established.

9. ATOMISTIC STRUCTURE OF MATTER

Up to now the quadratic action principle shows little relation to the classical results of general relativity. In general relativity the problem of matter appears in the following context. Any generally covariant Lagrangian has the following property. It contains the components g_{ik} of the metrical tensor. Varying with respect to the g^{ik}, we obtain a symmetric tensor whose divergence is zero. This tensor we call the "physical matter tensor." If for example L is identified with Maxwell's fundamental Lagrangian $-F_{ik}F^{ik}$, the variation with respect to g^{ik} yields Maxwell's "stress-tensor" S_{ik}, also called the "energy-momentum tensor" of the electromagnetic field. If matter is conceived in purely electric terms, this tensor in itself can be equated to the physical matter tensor. Generally L may be composed of further terms, in which case S_{ik} will be complemented by additional tensors. The sum of these tensors represents the physical matter tensor which we wish to denote by P_{ik}.

Einstein discovered the purely metrical origin of gravitational forces and established in the "metrical matter tensor" T_{ik} [see (5.18)] a symmetric tensor of second order whose divergence vanished identically. It could be derived variationally by varying the scalar Riemannian curvature R with respect to the g^{ik}. The equivalence principle demanded that *all forms of energy*

act also gravitationally. The mathematical formulation of this principle takes the following form:

$$T_{ik} = -\kappa P_{ik}. \tag{9.1}$$

It expresses the equilibrium between gravitational energy and all other forms of energy. The constant κ is a dimensioned scale factor by which energy density can be reduced to curvature. Its numerical value in the cm gram system is (the time being measured in natural units, i.e., c^{-1} sec):

$$\kappa = \frac{8\pi k}{c^2} = 1.863 \cdot 10^{-27} \text{ (cm/g)}. \tag{9.2}$$

From the standpoint of an action principle Einstein's fundamental equation should be written in the form

$$T_{ik} + \kappa P_{ik} = 0. \tag{9.3}$$

The significance of this equation is that the original physical Lagrangian L_0 is changed to

$$L = L_0 + (1/\kappa^{-1} R). \tag{9.4}$$

Moreover, the metrical tensor g_{ik} is added to the previous physical variables as additional variables of the field. However, such a purely external coupling of "metrical" and "physical" quantities never satisfied Einstein. He tried to find some geometrical interpretation of the "physical" part L_0 of the Lagrangian in order to conceive the *entire* L as one unified quantity.

Looking at our Lagrangian (8.10), deduced from an action principle which is quadratic in the curvature quantities but free of any external additions—except for quantities which are dictated by the mathematical nature of the problem—we are at first sight at a loss to see any relation to a Lagrangian of the form (9.4). Our Lagrangian contains the surplus variables p_{ik} which make a direct comparison difficult. We now make use of the fact, deduced earlier in Sec. 5 [see (5.16)], that in the absence of the potentials φ_i

$$p_{ik} = \lambda g_{ik} \tag{9.5}$$

is an *exact solution* of the field equations. At variance with the usual concepts we equate λ to an *excessively large* constant by considering it as the reciprocal square of a *length of subatomic dimensions*. Under these circumstances p_{ik} changes but little by the presence of the φ_i and we obtain a good approximation of our Lagrangian by considering (9.5) as an *exact law*, although in reality it holds only in approximation. We now introduce this solution into our Lagrangian (8.10) and thus reduce our Lagrangian problem from the original 10+10+4=24 variables to a problem of only 10+4=14 variables, *viz.* the 10 g_{ik} and the 4 φ_i. Now a direct comparison with Einstein's action principle becomes possible.

Our new Lagrangian is composed of a number of parts which we write down and analyze separately:

$$\begin{aligned} L_1 &= \lambda R \\ L_2 &= -2\lambda^2(1+4\sigma) \\ L_3 &= -\tfrac{1}{8} F_{ik} F^{ik} \\ L_4 &= -\tfrac{1}{2}(1+\sigma)(\varphi^\alpha{}_{,\alpha} - \tfrac{1}{2}\varphi^\alpha\varphi_\alpha)^2 \\ L_5 &= -\frac{\lambda}{2}(1+4\sigma)\varphi^\alpha\varphi_\alpha \\ L_6 &= \rho_i\left(\frac{\partial g^{\frac{1}{2}} g^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha} - \varphi^i\right). \end{aligned} \tag{9.6}$$

The first invariant appears originally in the form $\lambda B_{ik} g^{ik}$ but this differs from $\lambda R_{ik} g^{ik}$ by a divergence only and is thus replaceable by λR. Hence *Einstein's linear invariant comes into strong focus*. The largness of λ causes a practical *linearization* of our originally quadratic action principle, thus bringing it in harmony with demands of the equivalence principle.

The second invariant represents the traditional "cosmological term." If we wanted to adhere to the traditional view that the cosmological constant is of practically negligible dimensions, demanded only for cosmological purposes, we would choose σ as exceedingly near to $-\frac{1}{4}$, i.e. β as excessively large. This means an overwhelming emphasis of the invariant R^2 compared with the invariant $R_{ik}R^{ik}$. We are then back at the choice of Pauli and Weyl; [see (7.8)]. In this case we lose not only L_2 but also L_5. We deprive ourselves of an entirely essential building block for the construction of material particles of atomic size. Furthermore, we once more experience the above discussed difficulty concerning the free charge of an electrostatic field.

We depart from the traditional theory by assuming that the "effective cosmological constant"

$$\mu = -\lambda(1+4\sigma) \tag{9.7}$$

is not small but excessively large (although possibly much smaller than λ itself) by considering it as the reciprocal square of a length of atomic dimensions. (We assume μ to be positive which puts σ to the negative side of $-\frac{1}{4}$.) This means that the metrical substructure of the world, characterized by the equation [see (5.10)]

$$R_{ik} = -\mu g_{ik} \tag{9.8}$$

is very far from being flat, owing to the presence of a tremendous negative pressure which is uniformly present throughout the universe. The implications of this assumption will be discussed in the following section.

We now come to L_3 which is Maxwell's invariant. The standard theory which recognizes electricity and gravity as the two fundamental phenomena of nature, operates with L_1 and L_3 only. As we know, no stable particle can be constructed on this basis. But in our present considerations we have also L_4 and L_5 at our disposal. What kind of forces are generated by these

invariants? Varying with respect to g^{ik}, we obtain a tensor of the following form:

$$\Theta_{ik}=\alpha(\varphi_i\varphi_k-\tfrac{1}{2}\varphi^\alpha\varphi_\alpha g_{ik}). \qquad (9.9)$$

Interpreting φ_i as a velocity field[23] we recognize in this matter tensor the action of inertial masses, coupled with a pressure which is $\frac{1}{2}$ of the matter density. The forces thus generated are of a *mechanical* nature. We can thus answer the age-old puzzle: what kind of forces balance the tremendous electrostatic repulsion inside the electron which should lead to its explosion? The answer is that these forces are of a *mechanical* kind. Let us write down the equation which determines the vector potential [see (8.12)]:

$$\frac{1}{2}\frac{\partial g^{\frac{1}{2}}F^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}+[\mu+(1+\sigma)Q]\varphi^i+(1+\sigma)Q_{,\alpha}g^{\alpha i}=0. \qquad (9.10)$$

Let us examine this equation from the standpoint of static and spherically symmetric solutions, paying particular attention to the equation $i=4$ which determines the "scalar potential" φ_4. The scalar Q becomes negligibly small in weak fields (i.e., small φ_i); near the center $r=0$, however, it becomes very large. Moreover, we can conceive the equation (9.10) for $i=4$ as a homogeneous linear differential equation for φ_4 which contains, however, an *adjustable constant* because of the amplitude factor of Q which is freely disposable. We obtain a regular *eigenvalue problem* of a nonlinear type.[24] Apart from the realm of very small r we obtain the solution

$$\varphi_4=e^{-(2\mu)^{\frac{1}{2}}r}/r, \qquad (9.11)$$

and see that instead of the Coulomb potential we obtain a potential of the Yukawa kind.[25] But the singularity at the origin is avoided, due to the cooperation of the scalar Q which puts a very strong *negative* mass near the center of the particle, thus balancing the strong forces at the center.

We come to the conclusion that we are able to construct a *spherically symmetric and static electric particle which has no singularity anywhere.* The physical matter tensor generated by this particle has an entirely *insular* character. Matter is concentrated in little lumps of very small extension and a direct interaction between matter and matter is not possible. The Lorentz force is traditionally derived from the Maxwellian matter tensor, on the basis of the conservation laws. Hence it represents a matter-matter interaction. In the present theory an influence from particle to particle can only occur through *metrical* signals. Hence the electromagnetic vector potential must be of *metrical* origin and the Lorentz force must represent a matter-metric and not a matter-matter interaction.

This is foreshadowed by (8.13) which shows that the vector φ_i cannot be of the nature of a generating function since it is itself obtained from other field variables by differentiation. In fact this equation closely resembles the equation of Lorentz which connects the electric field quantities with the electric current vector:

$$\frac{\partial g^{\frac{1}{2}}F^{i\alpha}}{g^{\frac{1}{2}}\partial x_\alpha}=\rho^i. \qquad (9.12)$$

The difference is that g^{ik} is a symmetric, F^{ik} an antisymmetric tensor. The analogy goes still deeper. The significance of (8.13) is that we normalize our reference system in a suitable way. This normalization will involve a proper coordinate transformation. Assuming that the φ_i are small, the transformation assumes an *infinitesimal* character. Such a transformation is characterized by the law

$$\bar{g}_{ik}=g_{ik}-\epsilon(A_{i,k}+A_{k,i}) \qquad (9.13)$$

where ϵ is a small parameter. Thus the essential change demanded by our reorientation is that the role of the customary antisymmetric combination $F_{ik}=A_{i,k}-A_{k,i}$ is taken over by the symmetric combination

$$G_{ik}=A_{i,k}+A_{k,i}. \qquad (9.14)$$

The customary "nabla equation" between vector potential and current vector remains essentially unchanged, except that the continuity equation for the current vector is not a consequence of the connecting equation since the left side of the equation is free of any identities.

This reclassification of electric quantities within the framework of general relativity leads to surprising results. The vector φ_i, previously treated as the electromagnetic vector potential, becomes in actual fact the vector of the *electric current*. It is surprising to see that it is not the vector potential A_i but the electric current φ_i which has a direct influence on the matter tensor by replacing the vector potential in the Maxwellian stress tensor. This tensor is complemented, however, by additional mechanical terms. In the theory of Lorentz the electric current is a purely *extraneous* quantity, foisted on the field without any organic relation to the basic field quantities. It is not more than an "asylum ignorantiae," in the words of Einstein. In the present theory the electric current becomes a basic field quantity, organically related to the field and determined by an eigenvalue problem. The electron, before a "stranger of electrodynamics" (Einstein[26]), appears

[23] This is strictly speaking incorrect, because even in a static field the components φ_i $(i=1,2,3)$ are not zero. Since, however, the particle is small and we can average over the φ_i, the *average values* of the space components vanish and what remains is φ_4 which is tangential to the world line of the particle.

[24] The author's computations are not yet completed and will be reported at another occasion. The symmetry of the basic Lagrangian with respect to $\pm\varphi_4$ holds only if the metrical action of the φ_i is neglected. The mass-inequality of positive and negative electricity is not outside the scope of the present theory.

[25] See reference 5, p. 45.

[26] See A. Sommerfeld, *Electrodynamics* (Academic Press, Inc., New York, 1952), p. 236.

with a well-defined structure in which all infinities are avoided.

The reorientation of the electromagnetic quantities that is here advocated leads to a modified formulation of the dynamical law of the electron. The equation "time rate of change of momentum equals moving force" is retained but the "moving force" cannot be the Lorentz force since the antisymmetric combination $A_{i,k}-A_{k,i}$ has no significance. Its place is taken by the symmetric combination (9.14) according to the law

$$\frac{dp_i}{ds}=F_i=e(A_{\alpha,i}+A_{i,\alpha})\frac{dx^\alpha}{ds}, \qquad (9.15)$$

where A_i is the vector potential of the external field. The "momentum" of the electrically charged particle is defined as follows:

$$p_i=m_0\frac{dx^\alpha}{ds}g_{\alpha i}+2eA_i. \qquad (9.16)$$

If the correction term is transferred from the left to the right, the sum $A_{\alpha,i}+A_{i,\alpha}$ changes to the difference $A_{\alpha,i}-A_{i,\alpha}$ and the Lorentz force is once more restored.

The field momentum (9.16) differs from the customary expression[27] by the factor 2. The celebrated Einstein–de Haas e/m experiment which directly measures the correction to which the mechanical momentum is subjected in an external magnetic field, gave twice the expected value, in agreement with the law (9.16).[28]

10. METRICAL SUBSTRUCTURE

The most radical, though inevitable, departure of the present theory from the traditional views lies in the assumption that the material particles represent not more than a weak superstructure on a metrical substructure which is exceedingly strong. It is characterized by the equation

$$R_{ik}=-\mu g_{ik}, \qquad (10.1)$$

where μ is an exceedingly large constant. We have here a development somewhat analogous to Einstein's famous theoretical prediction[29] in 1907 that the traditional "kinetic energy" of Newtonian physics is not more than a very small modification of an exceedingly large energy source of the amount m_0c^2, connected with the very existence of the mass m_0. This enormous energy-source remains latent under ordinary circumstances because our measurements involve energy *differences* only and under such conditions the rest-mass usually drops out, due to the great stability of the atomic building blocks of the universe.

Equation (10.1) similarly assumes that all physical events are only small modifications of a fundamental structure which could be compared to a huge "rest-mass" of the universe. But is such an assumption not preposterous when we know that our ordinary geometry is so nearly Euclidian? Would not a law of the form (10.1) lead to a tremendous *explosion* of the universe? In actual fact this would only happen if the matter maintaining this tremendous curvature would act in a *coherent* way. But we can conceive the picture of a gas of very high temperature whose molecules fly around in all directions with practically light velocity and which create a huge uniform pressure, apparently due to the action of some very large force, but in actual fact caused by the statistical kinetic action of a large number of molecules. An explosion of tremendous power would occur only if by some miracle all the molecules could be directed into parallel paths.

We carry over this statistical picture in order to demonstrate that a metrical substructure of the form (10.1) is actually feasible, in spite of the apparently almost Euclidian character of our metric. We utilize a statistical device used by Lorentz to show that the macroscopic equations of Maxwell can be explained on the basis of electron theory, if we allow the statistical interaction of a huge number of electric charges in a region which is macroscopically small but yet large compared with the dimensions of atoms and molecules. Using a similar picture we want to calculate the *average values* of the curvature components, assuming that the metric is far from stationary. In fact, we want to assume that there exists a metrical radiation which is composed of exceedingly high frequencies and which propagates irregularly in every direction. We will now cut out a four-dimensional cube at some point of the world and evaluate the average values of the curvature components within the cube. The edge of this cube is chosen large compared with the wavelengths of the radiation. These wavelengths are so small that the cube can be of subatomic size.

For present purposes we depart from the previous real line-element of the signature $(-1, -1, -1, +1)$ and introduce Minkowski's Euclidian line-element $g_{ik}=\delta_{ik}$, considering $x_4=it$ as an imaginary variable. We consider a metrical tensor which differs from the Euclidian values by small amounts only

$$g_{ik}=\delta_{ik}+\gamma_{ik}. \qquad (10.2)$$

However, although the deviations themselves are small, the Γ quantities can become *very large* because the γ_{ik} are composed of a spectrum of very high frequencies. We start with Riemann's curvature tensor R_{imkn} which can be written in terms of an auxiliary tensor A_{ikmn} defined as follows:

$$R_{imkn}=A_{ikmn}-A_{inmk} \qquad (10.3)$$

$$A_{ikmn}=\tfrac{1}{2}\left(\frac{\partial^2 g_{ik}}{\partial x_m\partial x_n}+\frac{\partial^2 g_{mn}}{\partial x_i\partial x_k}\right)+\begin{bmatrix}ik\\ \alpha\end{bmatrix}\begin{bmatrix}mn\\ \beta\end{bmatrix}g^{\alpha\beta}. \qquad (10.4)$$

---

[27] See, e.g., H. Goldstein, *Classical Mechanics* (Addison-Wesley Press, Cambridge, 1951), p. 49.

[28] The customary explanation refers the discrepancy to the "spin" of the electron. The spin action of the electron is deeply interwoven with the here discussed re-evaluation of electric quantities. The deeper analysis of this problem transcends, however, the limitations of the present paper.

[29] Albert Einstein, Jahr. Radioakt. 4, 411 (1907).

The terms which are derivatives, cannot contribute to the average values since the derivative of a periodic function is once more periodic and its average value zero. It is sufficient to consider the last term of (10.4), and this again can be written with practically sufficient accuracy in the form

$$a_{ikmn}=\begin{bmatrix} ik \\ \alpha \end{bmatrix}\begin{bmatrix} mn \\ \alpha \end{bmatrix}. \qquad (10.5)$$

We now make the statistical assumption that the various components of the tensor quantity

$$\Gamma_{ik,\,m}=\begin{bmatrix} ik \\ m \end{bmatrix} \qquad (10.6)$$

are statistically uncorrelated, so taking the average values of products involving two different components gives zero. Average values of the squares of components cannot give zero since the average of both $\sin^2\omega x$ and $\cos^2\omega x$ is $\frac{1}{2}$. (In a purely Euclidian world these averages would be composed of a sum of squares and be necessarily positive. In a Minkowskian world some of the terms become negative, because of the differentiation with respect to the imaginary x_4.) In consequence of this statistical behavior average values of a_{ikmn} will become proportional to $\delta_{im}\delta_{kn}+\delta_{in}\delta_{km}$. We assume that the factor of proportionality is a universal constant C:

$$\bar{a}_{ikmn}=C(\delta_{im}\delta_{kn}+\delta_{in}\delta_{km}). \qquad (10.7)$$

This gives

$$\bar{R}_{imkn}=C(\delta_{in}\delta_{km}-\delta_{ik}\delta_{mn}). \qquad (10.8)$$

The constant C is of the order of magnitude $(\epsilon\omega)^2$ and can become very large, in spite of the smallness of ϵ. The inhabitants of this world, who cannot measure local values but only averages—since the elementary particles which they can use as measuring rods are themselves large compared with the basic wavelength—will find that their metric is nearly Euclidian and yet the curvature of their world uniform but exceedingly high. This would not be possible in terms of local values, of course. However, the *contracted* curvature tensor R_{ik} may come out as proportional to g_{ik} (with a very high proportionality factor) even in the local sense. Hence, it is not absurd to assume that the substructure of the world is characterized by the law (10.1) and yet the g_{ik} are of the nature of constants, modified only by oscillatory variations of exceedingly high frequencies but small amplitudes.[30]

11. SUMMARY

The present paper investigates the possibilities of a Riemannian geometry characterized by an action principle which is not linear in the curvature components, as the choice made by Einstein, but quadratic. Such a Lagrangian is advocated by the property that the basic action integral remains invariant not only with respect to arbitrary coordinate transformations but also with respect to an arbitrary choice of the scale in which lengths are measured. Hermann Weyl advocated such an action principle at an earlier date, but in connection with a proposed generalization of Riemann's geometry. The author's aim—spreading over a span of more than 25 years—was to draw all the possible logical conclusions from the quadratic action principle, without abandoning Riemann's geometry.

Mathematical analysis of this problem was carried out by a new mathematical method, based on the method of the Lagrangian multiplier, by which the problem is reduced to a set of field equations of first order in an increased number of variables. This demands 100 variables. It seemed more convenient, however, to admit differential operators of second order and formulate the problem in 20 variables. The results can be summarized as follows.

(1) Einstein's field equations $R_{ik}=0$ are included among the solutions but so are the more general "cosmological equations" $R_{ik}=\lambda g_{ik}$. The merit of the latter solution is that it introduces a dimensioned quantity λ which normalizes the lengths of the universe. To assume that the normal length of the universe is of cosmic magnitude means that we miss the opportunity for the explanation of atomism. The present theory assumes that the fundamental unit of length is of subatomic dimensions. This makes the cosmological constant exceedingly large which means that the substructure of the world is not flat but highly curved.

(2) Weyl's generalization of Riemann's geometry for the purpose of introducing the vector potential turns out to be unnecessary since the vector potential appears in the action principle quite naturally as the result of a canonical transformation. Later, a purely metrical interpretation of the vector potential is found, on the basis of splitting the contracted curvature tensor R_{ik} into a simplified operator B_{ik}, which is free of all identities, and a second part which involves the symmetrized gradient of a coordinate-dependent vectorial quantity V^i. It is found that this vector, in the ordinary gravitational theory normalized to zero, should actually be equated to the vector potential φ^i.

(3) In view of the largeness of λ we succeed with an approximate linearization of the action principle and a direct comparison with classical relativity becomes possible. A deeper analysis of the determining equation of φ_i reveals that φ_i was wrongly interpreted as the vector potential of the electromagnetic field. It is in actual fact the electric *current vector*. At variance to the theory of Lorentz in which the current vector is foisted on the field as an extraneous quantity without definite structure, the vector φ_i is determined by a differential equation of second order which possesses

[30] For another approach to the problem of the substructure, based on a Fourier analysis of the metrical manifold as a whole, see the author's earlier paper "Matter waves and electricity," reference 3, II, where the idea of a cosmological constant of subatomic dimensions was discussed for the first time.

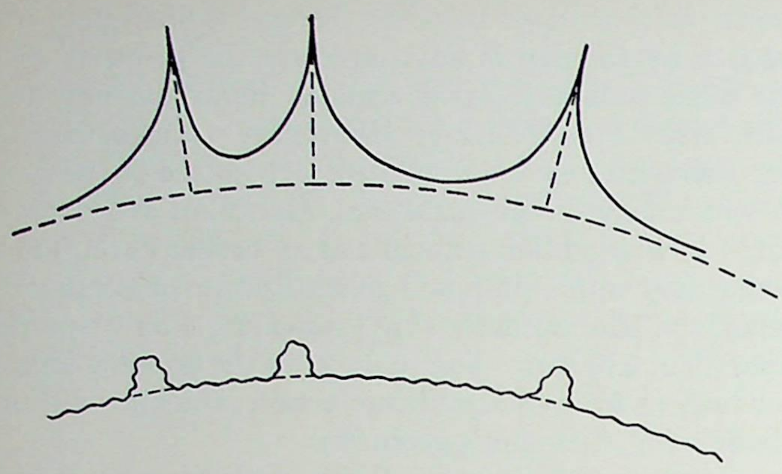


FIG. 1.

static and spherically symmetric eigensolutions. A static and stable electron can thus be constructed, free of all infinities. The forces which counteract the strong repulsion of the electrostatic forces, are of a mechanical kind.

(4) The matter tensor is reduced to very small portions of space and does not have that diffused character that the Maxwellian stress tensor attributed to it. The role of the vector potential in the Maxwellian tensor is taken over by the current vector. Matter becomes strictly insular and the dynamical influence from particle to particle occurs purely through the medium of the metric. The Lorentz force loses its primary significance and comes about by the interaction of two terms on opposite sides of the equation of motion, the one belonging to the momentum of the particle, the other containing the symmetric combination $A_{i,k}+A_{k,i}$, instead of the traditional antisymmetric combination $A_{i,k}-A_{k,i}$, which in the present theory has no primary significance.

(5) The superstructure of the metric in the form of material particles is a weak modification of the substructure which represents a tremendous reservoir of momentum and energy in potential form. That under ordinary circumstances the weak modification becomes of paramount importance and the much stronger substructure remains latent, finds its explanation in the fact that the superstructure is of a *static* character while the substructure represents a statistically distributed radiation field of exceedingly high frequency which is comparable to a capricious sequence of constant seismographical "tremors," too small to be observable under ordinary macroscopic circumstances but of paramount importance for the question of radiation, i.e. the interaction of particle and field. It seems thus justified to call the metrical substructure, composed of a seething cauldron of ultra-high-frequency oscillations. the "tremorfield." Figure 1 gives a schematic illustration of the general modification of ideas that the present theory suggests, compared with the traditional concepts. In the traditional view we have a smooth background, almost completely flat, except for a very slight curvature, demanded by the cosmologically closed nature of the universe. Erected on this background we find the material particles, growing to infinity near the center (or at the center) and decreasing in strength toward the periphery, but the decrease is mild enough to make a direct matter-matter interaction possible; (the Lorentz force obtained from the conservation law of the matter tensor). The second picture illustrates a highly agitated "rippled" background and the material particles grow out of this agitated background as little humps, completely isolated from each other. The signals emanating from a particle and reaching another one are of a purely metrical nature and belong thus to the realm of gravitational action. The motion of a particle takes place due to the response of the particle to the external metric which does not occur according to the geodesic principle.

The tremendous complication which comes into the picture due to the presence of the substructure demonstrates that our conceptual understanding of physical action cannot be brought down to a simple and elementary level. The problem of radiation is much more than a secondary interaction between field and particle, induced by the energy loss accompanying the accelerated motion of a particle. The momentum-energy exchange between substructure and superstructure opens a new perspective for the deeper understanding of the mysterious quantum phenomena and it seems possible that Heisenberg's uncertainty principle is deeply interwoven with the statistical nature of the tremorfield. To find the solution of these problems on the basis of the fundamental field equations may be far in the future. The author's aim in this investigation was only to outline a new solution of the problem of general relativity which may finally bring gravity, electricity and quantum physics into one unified structure, based on the immortal thought constructions of Albert Einstein.

Papers from the Conference on the Role of Gravitation in Physics Held at the University of North Carolina, Chapel Hill, North Carolina, January 18–23, 1957

Introductory Note

THE following papers were prepared in connection with the Conference on the Role of Gravitation in Physics which was held in Chapel Hill, January 18–23, 1957. The conference was initiated by the North Carolina Project of the Institute of Field Physics, Inc., established in 1956 in the Department of Physics of the University of North Carolina, Chapel Hill. The sponsors of the conference were the International Union of Pure and Applied Science (with a financial contribution from UNESCO), the National Science Foundation, the Wright Air Development Center, and the Office of Ordnance Research. The conference was organized—with the collaboration of the Institute of Natural Science of the University of North Carolina—by the following steering committee: F. J. Belinfante, Purdue University; P. G. Bergmann, Syracuse University; B. S. DeWitt, University of North Carolina; Cecile M. DeWitt, University of North Carolina; F. J. Dyson, Institute for Advanced Study; and J. A. Wheeler, Princeton University. It is a pleasure to mention the interest shown in the conference by the officers of the University, by the Governor of the State of North Carolina, who welcomed the physicists to Chapel Hill at a luncheon in the Morehead Planetarium, and by all who have made the Chapel Hill conference possible and have helped in its organization.

The conference was planned as a working session to discuss problems in the theory of gravitation which recently have received attention. A report of the proceedings of the conference can be obtained from Wright Air Development Center, Wright-Patterson Air Force Base, Ohio. The WADC report endeavors to record the discussions, confusions, opinions, and hopes expressed during the conference. A separate need was felt to assemble, in an issue of a scientific journal, the papers on gravitation which were ready for publication—whether or not they had been presented at the conference. Participants in the conference are grateful to the editors of *Reviews of Modern Physics* for making available the space necessary for these papers.

This collection of papers cannot strictly be considered as a review of present knowledge about gravitation. A noticeable increase of interest and activity in the theory of gravitation and related matters has taken place only recently, following a period of relative quiet. The problems attacked have not yet settled down so that they cannot be viewed with any kind of real perspective. At present the most that can be hoped for is that these papers will give a fairly reasonable picture of current activity in the field of relativity physics and stimulate greater awareness of a whole series of fundamental questions which have yet to be answered.

BRYCE S. DEWITT

Summary of the Chapel Hill Conference*

PETER G. BERGMANN

Department of Physics, Syracuse University, Syracuse 10, New York

This conference may be summarized under two main headings, classical theory of gravitation and quantization of general-relativistic theories; a third topic, the theory of measurement, forms a link between these two topics. I shall not attempt to summarize the discussion of cosmological problems by Dr. T. Gold and Dr. A. Lilley, because cosmology is a field of its own and, at least at present, not intimately connected with the other aspects of general relativity to which this conference has been devoted.

I. NONQUANTUM THEORY OF GRAVITATION

FOLLOWING roughly the outline of the conference itself, I shall begin with experimental work, past, current, and proposed. Professor J. A. Wheeler has given us a summary of the present status of the so-called "critical tests," that is the precession of planetary orbits, the deflection of light passing close to the sun, and the gravitational red shift. I think as nonspecialists we may consider the situation in this respect quite satisfactory, but additional work, leading to higher precision, is highly desirable. A very different type of experimental test of general relativity forms the first topic of Professor Dicke's talk. Whereas at least the two first of the critical tests are designed to provide confirmation of the detailed quantitative predictions of Einstein's 1916 theory of gravitation, the experiment by Eötvös tests the principle of equivalence, the root of *any* general-relativistic theory. I think that we shall all await with anticipation the outcome of these experiments by Dicke.

The remaining experiments suggested by Professor Dicke, which are not yet in the stage of actual planning, are mostly designed to see whether there is some minute variability in the universal parameters of elementary particles, which we used to consider as constants. I do not believe that the theories sponsored for instance by P. Jordan have as yet been clarified conceptually to the extent that they merit experimental testing. In this connection I call attention to a recent paper by M. Fierz in *Helvetica Physica Acta.* Of course, it might be interesting to test for variability of atomic constants irrespective of any detailed theory. Clearly a positive result of such experiments or observations would have most serious implications both for cosmology and for the theory of elementary particles.

Let me take up as the next topic the mathematical investigation of general-relativistic theories that have been reported principally by the members of the French group. I believe that the techniques pioneered by A. Lichnérowicz are applicable not only to the general theory of relativity and to the asymmetric theory, but to general-relativistic theories in general. The law of propagation for infinitesimal disturbances of the field will define the light cone in a physically convincing manner. I am very gratified to learn that there exists a method for formulating the Cauchy problem, but I take it that we do not as yet possess a method for constructing the variables whose values at time t are determined uniquely by the initial conditions set at the time t_0. If we had such a construction, these variables would presumably be identical with the "true observables," about which I shall have more to say later.

None of us can fail to be impressed with the availability of at least a few global theorems in general relativity, though one might wish that more, and more powerful ones, may be discovered in the years to come.

I shall now turn to the investigations that have been concerned with specific physical models that can be, or ought to be, constructed with the help of Einstein's field equations. It appears to me that the geons, which have been invented by J. A. Wheeler and which have been discussed here by him and by members of his group, cannot yet be considered to be actual solutions of the field equations, with or without electromagnetic terms. Geons are obtained in a manner that might be called a self-consistent approach: a collection of particles or waves, which is described in terms of statistical assemblies, with random phasing, serves as the source of a static gravitational field; the latter in turn confines the constituent particles to a limited region in space. Though these models undoubtedly have considerable heuristic value, I believe we all would be happier if it could be established that there exist at least some rigorous solutions of the field equations with properties similar to the geons.

In my opinion the most important nonquantum problem that has been discussed at this conference is the existence of gravitational waves. Actually there is a whole series of questions to which we should like to have answers. Many years ago A. Einstein and N. Rosen constructed rigorous waves radiated off the axis of a cylinder. According to N. Rosen and according to J.

* Work by the author in this field is currently being supported by the National Science Foundation and by the Air Force Office of Scientific Research. In this paper I have consolidated the introduction to the sessions on the quantum theory of gravitation with my summary talk. Of course I have been influenced by the talks and discussions of the conference; but I must accept personal responsibility for the selection of material and for all opinions voiced.

Weber these waves do not carry energy in any physical sense. It may well be that because of the line singularity extending into spatial infinity any attempt to balance energy, in the sense of an equation of continuity, is foredoomed to failure. In that case we should want to focus our attention on spherical waves. At present we do not know whether there exist any rigorous solutions of the field equations that may be interpreted as spherical waves, nor do we know whether such solutions carry energy from a center of radiation toward infinity.

Another unsolved question is whether a first approximation of a double-star system gives rise to gravitational waves and whether these waves carry energy proportional to the square of the amplitude. From the point of view of the astronomer, all these questions may appear highly academic, because gravitational waves, if generated by astronomical systems, undoubtedly play a negligible role in the energy budget of stars, compared to electromagnetic radiation and compared to the exchange of particles with the surrounding space. But in view of our interest in the role that gravitation, and particularly its quantum properties, may play in microphysics, the existence and the properties of gravitational waves represent an issue of preeminent physical significance.

Only in passing I should like to remind you of the investigations by the Princeton group of the stability of model universes, and of the pair of accelerated mass dipoles constructed by H. Bondi. The latter is one of the very few rigorous solutions of the field equations that are not static; as a model it may well prove of great value in future investigations.

I shall comment on the wormhole picture presented by Professor J. A. Wheeler in my final remarks on elementary particle problems.

We have had some discussions concerning the prospects of unified field theories and a number of different views have been presented. Wheeler considers that it may well be that all the physical fields known to us at present will turn out to be manifestations of but one field, the original gravitational field of Einstein's theory of 1916. Others are following up the possibilities inherent in asymmetric field theories. A third approach, unfortunately not represented at this conference, is being followed by Professor O. Klein; he has been interested in five-dimensional manifolds with periodicity properties such that there exists a sort of isotopic spin transformation. Personally, I am convinced of the long-range desirability of some unified field theory; but I have the impression that right now we are obtaining so much information about new fields and new types of particles that it may be advisable to wait for a few years before making a new determined attempt at unification.

II. THEORY OF MEASUREMENT

In recent years several different workers have asked which physical aspects of general-relativistic fields are susceptible to observation and measurement. This problem differs from the corresponding problem in Lorentz-covariant theories in that in the latter type of theory a world point is identifiable by means of its coordinates, and irrespective of the physical fields present. Accordingly we may, at least classically, speak of the value of a field, or a field component, at such a world point unambiguously. But in general relativity the identification of a world point in terms of its coordinate values lacks any invariant meaning unless the coordinate system is fixed in its immediate vicinity by means of physically real instrument components, such as rods and clocks. Naturally the presence of such instruments will in turn affect the field to be measured at that world point.

Classically we may make the instruments negligibly small and light and thus minimize the reaction. Even so I think it represents real progress that Dr. Pirani has shown us how to measure (classically) the individual components of the Riemann-Christoffel curvature tensor.

As for the quantum theory of measurement we have the contributions by Dr. H. Salecker and the remarks by Professor L. Rosenfeld. I believe that in conceptual experiments it is dangerous to introduce for the purpose of observations instruments that are not made an integral part of the physical system itself. The experiment with a diffraction grating suggested by Dr. Salecker can however be freed of this objection. Suppose we consider a small planetary system with gravitational attraction in a low quantum state. Even if the central body is assumed to possess a very large mass, the DeBroglie wavelength of the small particle depends on its mass. Hence the quantum orbits will depend on the planetary mass as well, and we find that such a model is not subject to an equivalence principle. This result is not to be construed as a violation of the usual principle of equivalence in quantum theory: properly speaking the principle of equivalence refers to local conditions, whereas the phase integral rule applies to a global model.

If I understand Professor Rosenfeld's remarks rightly, he suggests that the safest procedure is to have the observability of a physical quantity decided on the basis of formal quantum theory. Subsequently the edicts of the theory can be buttressed by conceptual experiments.

III. QUANTUM THEORY OF GRAVITATION

Basically we feel the need for a quantum theory of the gravitational field, because we cannot reconcile ourselves to the idea that classical and quantum fields can exist side by side. More particularly it appears unlikely that a mass point that gives rise to a classical Schwarzschild field is itself subject to Heisenberg's uncertainty relations. This, I believe, is the principal argument in favor of quantization. On the positive side, we may perhaps hope with Pauli that quantizing the metric tensor may ameliorate the infinities of the other field

propagators along the light cone and that it may make a contribution to the theory of elementary particles.

The task of quantizing the metric field is not yet really close to its solution. If I am to describe its present status, I would divide all approaches known to me into those that take their point of departure from some canonical or Lagrangian formulation of the theory and those that attempt a path-integral quantization. There appears little doubt that for all approaches belonging to the first class the construction of so-called *true observables* is an unavoidable preliminary.

For those who are not familiar with the terminology let me repeat that a true observable is a physical variable whose value is independent of the choice of coordinate system, gauge frame, and the like. Incidentally, true observables are also the appropriate variables for the formulation of the Cauchy problem. Analytically they may be described as variables whose Poisson brackets with all the constraints of the theory vanish.

J. Géhéniau, and independently A. Komar, have proposed to construct true observables with the help of methods of differential geometry. Unfortunately we have been unable to learn of this work directly from them at this conference. E. Newman and I have worked out a method by which we obtain true observables in the course of a systematic approximation procedure; our work may be suspect because we have no assurance that our approximation converges toward the correct quantities. Finally one might hope that the method used so successfully by Fermi in electromagnetic theory will lead toward the true observables in the theory of gravitation; A. Janis and I have looked into this question and have reluctantly concluded that this hope is probably unjustified.

I come now to the proposals involving quantization by means of path integrals. Within this classification there appear two schools of thought, those who believe that true observables are a prerequisite for this method of quantization as well, and those who propose to go ahead and calculate formal expressions in terms of the original $g_{\mu\nu}$, in the hope that these expressions will automatically be useful propagators. Feynman has given this latter group encouragement. Perhaps just to balance any effect that his words may have I should like to voice my opinion to the effect that in the theory of gravitation any calculation that you may attempt will involve such enormous amounts of labor, and will lead to results of such complexity, that generally a look at the final expressions will not enlighten us much.

In conclusion I should like to comment, briefly at least, on the word pictures of elementary particle theory presented to us by Professor J. A. Wheeler and Dr. S. Deser. If we permit the components of the metric tensor to fluctuate widely and without regard to each other, there will be many regions in which the signature of the metric changes and which no longer correspond to a singly connected Riemannian manifold. Thus Wheeler arrives at a figure of 10^{60} wormholes in the volume of space filled by a single electron. According to this picture, elementary particles are by no means simple structures but some sort of collective modes of the froth of the vacuum. Deser proposes a different approach. He wants to carry out a path-integral calculation with a Lagrangian combining the gravitational, electromagnetic, and other fields. He proposes to proceed first with a fixed metric tensor, so that he may quantize the other fields in a given c-number Riemannian manifold; the quantization of the metric field itself is to be left for the final stage of the calculation. Feynman has attacked Deser's proposal on the grounds that it fails to resolve the divergences of the self-energy diagrams, but I cannot accept Feynman's argument as conclusive. At any rate both Wheeler's and Deser's proposals promise a fresh attack on the structure of elementary particles. It will be very interesting to see whether either of the proposals can be made to yield results.

In summary then, I believe that in the time between this conference and the next relativity conference (planned in Europe for the summer of 1958) we have a good chance to make significant progress in the two classical problems concerning gravitational waves and true observables, and that thereby we may also contribute to the task of quantizing general-relativistic fields.

Principle of Equivalence and the Weak Interactions

R. H. DICKE
Palmer Physical Laboratory, Princeton University

AFTER 40 years the Einstein theory of gravitation still rests on little more than the experimental and observational evidence available when the theory was first constructed. This consists of the very precise astronomical observations on planetary motion, including the small departures from the results of Newtonian theory in the case of the orbit of Mercury, and the experiments on the equivalence of gravitational and inertial masses.

Shortly after the introduction of general relativity, the gravitational deflection of light was found to be in agreement with predictions with an accuracy of about 10%. More recent observations have tended to give deflections too large by about 10% with about the same probable error.[1]

The gravitational red shift could be expected from elementary considerations of energy conservation involving photons, and hence does not represent a crucial test of General Relativity. The red shift observed for the white dwarf stars is in rough agreement with expectations. For the light from the sun, there is a strange anomaly that has not yet been satisfactorily explained.[2]

After the advent of general relativity, fundamental experimental work on gravity essentially ceased. Only one significant new result needs to be mentioned.[2a] In connection with the practical problem of locating oil, sensitive gravimeters were devised. These have been improved to the point where they show tidal variations in the gravitational acceleration.[3] This work is of fundamental significance as the agreement with expected tidal variations is sufficiently good to rule out an ether drift effect such that the acceleration of gravity would vary with angle of acceleration relative to the direction of motion of the earth relative to galactic matter. An effect of the order of $(v/c)^2$ can be pretty well excluded, although observations have not yet been extended over a whole year.

Newtonian mechanics is even adequate for the construction of a cosmology in agreement with the observations.[4] Hence, no strong support for general relativity can be obtained from the present state of galactic observations. This conceivably could improve greatly in the next few years.

EÖTVÖS EXPERIMENT AND THE PRINCIPLE OF EQUIVALENCE

The Eötvös experiment[5] shows that, with an accuracy of about $1/10^8$, the inertial and gravitational measures of mass are equivalent, independent of material, at least for the exotic materials tested by Eötvös which included tallow and snakewood. It would be a serious mistake to assume, as is sometimes done, that this proves that the principle of equivalence is exact or even that all deductions from the principle of equivalence should have an accuracy of $1/10^8$.

Actually, the great accuracy of this experiment allows far reaching conclusions to be drawn. For example, with an accuracy of about $1:10^7$ we can conclude that the ratio of weight to mass for the electron plus proton is the same as that of the neutron. With an accuracy of about $1:10^5$ we can conclude that the reduction in the mass of a nucleus resulting from nuclear binding forces is accompanied by a like reduction in weight. In similar manner with an accuracy of $1:10^5$ we conclude that the electrostatic reduction of the nuclear binding energy is accompanied by a like increase in weight. With an accuracy of about $5:10^{+3}$ the same conclusion can be drawn concerning the binding energy of the orbital electrons.[6]

From these facts we conclude that the strong interaction constants are at least approximately position independent. If the strong interactions were to vary with position, the ratio of binding energy to particle rest energy would depend upon position and there would be an anomalous contribution to the weight associated with this variation. This can be seen by considering a small displacement of an atom. If this were accompanied by a small change in the fine structure constant or the interaction constant associated with nuclear forces, work would be required to produce the resulting change in binding energy. This added work would imply an anomalous weight of the body if energy were to be conserved.

---

[1] Whittaker, *History of the Theories of Ether and Electricity* (Nelson, London, 1953) Vol. 2, p. 180.

[2] E. Finlay-Freundlich, Phil. Mag. 45, 303 (1954).

[2a] Although not a new result, it should be mentioned that the very careful observational work of G. M. Clemence [Revs. Modern Phys. 19, 361 (1947)] has resulted in considerable improvement in the accuracy of our knowledge of the perihelion rotation effect.

[3] B. Baars, Geophys. Prospecting 1, 82 (1953).

[4] H. Bondi, *Cosmology* (Cambridge University Press, Cambridge, England, 1952).

[5] R. v. Eötvös, Ann. Physik 68, 11 (1922).

[6] See A. H. Wapson and G. J. Nijgh, Physica 21, 796 (1955), for a similar analysis.

TABLE I. Fundamental constants written as dimensionless numbers. Physical and astrophysical constants (units of $\hbar$, c, m).

10^{0}	10^{20}	10^{40}	10^{80}
Masses elementary particles $e^2=\alpha\hbar c$ Other "strong" Coupling const^{+2}	Reciprocal "weak" coupling const^2 β decay, μ decay, $\pi\rightarrow\mu$ decay, etc., etc., etc.	Reciprocal gravitational coupling const^2 $\left(\text{Ratio } \dfrac{\text{electrical force}}{\text{gravitational}}\right)$ Age of universe Hubble radius of universe	Number of particles in universe to Hubble radius

The weak interactions (Fermi interactions and gravitation) do not contribute sufficiently to the binding energy of a small complex system to allow any conclusion to be drawn regarding their constancy. Consequently, the Eötvös experiment in itself does not preclude variation of the weak interaction constants with position.

Normally it would seen reasonable to assume that constancy of strong interactions could be inferred to hold as well for weak interactions. Stated in other words lack of experimental information concerning constancy of weak interactions should not be taken as evidence that the interactions are not constant.

On the other hand, there are independent reasons for believing that there is something unusual about the weak interactions and that, in fact, these interaction constants vary with position and time. The line of reasoning is a bit tortuous and starts with a consideration of the sizes of the physical and astrophysical constants.

In Table I is tabulated a few representative examples of the fundamental constants of physics and astrophysics. They have been expressed as dimensionless numbers by introducing c, $\hbar$, and m, the mass of the electron, as units. By very great stretching of the meaning of "order of magnitude" the numbers can be said to all have the order of magnitude given in the first row of Table I.

Expressed in this way these constants have two remarkable properties. First, the large size of the reciprocal of the square of the weak interaction constants is unexpected. For, if nature is not capricious, such dimensionless numbers would be expected to result some day from theory, and the solutions of equations. Surely it is unreasonable to expect such a large number to result as a function of e, π, etc.

The second remarkable property of this table of numbers is the apparent interrelation between numbers, with the Fermi interaction constant seemingly related to the gravitational interaction constant and these weak interaction constants apparently related to the astrophysical constants.

Now what sense can be made of these systematics? There are three classes of explanations which have been offered in the past. The first is probably the one most commonly believed. According to this explanation, the apparent interrelations are not very close and there are only a few constants significantly different. Also, it is claimed that nature is somewhat capricious. A weak interaction is weak, and that is all there is to it.

The second type of explanation is of the type put forward by Eddington. This type of explanation seeks to obtain all of these dimensionless numbers, physical and astrophysical, as solutions to very involved equations whose origin is not always crystal clear.

A third class of explanations is associated with the names of Dirac[7] and Jordan.[7a] The apparent regularities between numbers are taken to indicate interconnections not presently understood. The age of the universe clearly varies with time. It is assumed that all numbers of the order of 10^{40} vary together and are, hence, proportional to the age of the universe. The number 10^{80} should vary as the square of the age of the universe.

The problem of the large size of these numbers now has a ready explanation. There is one large capricious or random number. The epoch of man would be expected to be a time which on an atomic scale is large and random. Man with all his complexity could not have evolved in a characteristic atomic time. Because of the interconnections, occurrence of a single large number of capricious origin is now sufficient to account for all the large numbers.

The characteristic features of this third class of explanations is the assumption of interrelations between constants and the occurrence of a single large number, statistical in origin.

This third type of explanation is very satisfactory in many ways, but appears to be incompatible with the strong version of the principle of equivalence usually assumed. According to Dirac the gravitational interaction constant should be getting weaker with cosmic time. In a frame of reference moving relative to the cosmological co-moving coordinate system, the cosmic time axis is tilted, and the gravitational interaction is a function of both time and position. As was discussed previously a space variation in the weak interaction constant would lead to a violation of the usual strong version of the principle of equivalence. The local gravitational acceleration now depends upon the pro-

[7] P. A. M. Dirac, Proc. Roy. Soc. (London) A165, 199 (1938).

[7a] P. Jordan, *Schwerkraft und Weltall, Grundlagen der Theor. Kosmologie. Braunschweig* (Vieweg and Sohn, 1952).

portion of weak interaction energy in the total binding energy. Thus, all bodies do not fall with the same acceleration and the principle of equivalence is satisfied only as an approximate relation.

To summarize, Dirac's explanation of the regularities in Table I is incompatible with the usual strong version of the principle of equivalence. On the other hand, if one is to believe that nature is not capricious, Dirac's explanation, or one of this general type, seems to be the best available.

As this represents all the direct experimental evidence concerning the constancy of the weak coupling constants, one must turn to indirect evidence, evidence concerning the distant past and available from geology, biology and astronomy.[7b] This is a very difficult task. There are enormous numbers of degrees of freedom in the earth and solar system. Projections into the past are heavily dependent upon what one assumes for physical laws in carrying out the projection. Consequently, it is often difficult to disentangle the actual observations from the interpretation, based as it is upon preconceived notions. This is particularly true of the qualitative aspects of the earth science, which is the main source of information concerning the past.

It seems very unlikely, in view of the complexities, that such indirect information can ever be used to prove that the weak interactions constants vary with time. For any effect that is observed there are a number of possible explanations.

On the other hand, the converse is not true. If the expected variation in the weak interaction constants were to lead to results incompatible with present observations, the hypothesis could be eliminated. It becomes important, therefore, to explore the implication of the assumptions that these constants vary with time.

In the interest of conciseness the conventional viewpoint is not presented here. It can be read in many of the references cited. Thus the interpretation presented here is to be viewed as a possible one but by no means the only one.

EARTH'S TEMPERATURE

Under the assumption that the gravitational constant were to vary with time, the earth's temperature would vary with time. This was first pointed out by Teller[8] who showed that the sun's radiation rate would vary as G^7 where G is the gravitational constant. Assuming conservation of angular momentum, the temperature of the earth was computed by Teller to vary as $G^{2.25}$. From this he concluded that if G varied inversely as the age of the universe, dinosaurs would have found it uncomfortably warm.

Since Teller's paper the galactic red shift data has given an age for the universe roughly 3 times that assumed by Teller and the effect is now not nearly as serious. Also, Teller did not take into account the marked atmospheric changes which would accompany a general increase in the solar constant.

The heat balance of the earth's atmosphere is very complicated and is affected by many factors; circulatory patterns, carbon dioxide content, water vapor content, and a host of others. Any extrapolation several billion years into the past is of necessity unreliable.

Ignoring the finer details, the most important effect of increased temperature upon the earth's atmosphere is the resulting increase in water vapor content. With a very moderate increase in the earth's mean temperature, the amount of atmospheric water vapor would increase greatly.

An increase in water vapor content would be expected to affect the radiation balance of the earth and to lead to either a lower or higher mean temperature for the surface of the earth. Thus the great temperature sensitivity of the water vapor pressure serves either to stabilize the earth's temperature or to make it unstable.

In order to decide between these two alternatives it is helpful to consider a simplified atmosphere consisting mostly of water vapor. A lower bound on the temperature of the atmosphere at any height can be obtained from the vapor pressure curve for water. The cloud cover of such an atmosphere would be essentially 100% as any hole in the cloud cover would lead to convection with resulting cloud formation. Convection is not of the usual type as the down-flow is almost completely liquid water with a water vapor up-flow.

It would appear that there are two effects tending to reduce the surface temperature of the earth. First, with 100% cloud cover the albedo of the earth is very high. Second, because of the small temperature gradient of a saturated vapor atmosphere, the temperature of the surface of the earth is only slightly greater than the temperature of the emitting layer of the atmosphere.

Assuming an age of the universe of 6.5×10^9 years and a solar energy flux at the earth varying as t^9, the solar flux at the earth 1.7 billion years ago would have been almost 15 times the present value. On the other hand, assuming a visual albedo for a cloud banked earth of 0.8, the surface temperature would have increased to 100°C only.

For an incident energy flux much greater than the above value, the regulatory mechanism fails to operate. This comes about in the following way: The temperature of the emitting layer is ultimately raised so high that the whole of the outer part of the atmosphere is at a temperature above the dew point. The cloud cover disappears and the convective mechanism becomes the normal type with vapor being heated by compression as it falls. Without precipitation the temperature gradient becomes that of adiabatic equilibrium. This leads to a greatly increased temperature at the earth's surface with more evaporation of water resulting. But the increased atmospheric pressure leads to a still greater

[7b] P. Jordan (reference 7a) has also discussed some of the geological implications of a varying gravitational interaction.
[8] E. Teller, Phys. Rev. 73, 801 (1948).

surface temperature with more evaporation. The atmosphere has become unstable and the whole content of the oceans evaporates producing an atmospheric pressure of about 3×10^8 dynes/cm^2 and a temperature greater than 1000°C at the earth's surface.

This high-temperature form of a water vapor atmosphere is opaque to the sun's radiation. However, the heat flow from the earth at a rate of 10^{-6} cal cm^{-2} sec^{-1} is sufficient to maintain convective equilibrium.

A temperature curve for the earth based upon the above model is plotted in Fig. 1 as curve *A*. It may be noted that for times in the past greater than 1.9 billion years the atmosphere has the high temperature form. This part of the curve is obtained by assuming an albedo of 0.6 and an emitting layer in the top 0.006 of the atmosphere. Also, because of the very small heat transfer in the lower part of the atmosphere, the temperature distribution is assumed to be adiabatic with the temperature at the "top" of the atmosphere being the lowest encountered over the whole of the earth's surface. This is assumed to be 100° below the average emitting temperature.

The low temperature curve is obtained by assuming that at the time -1.9 billion years the albedo of the earth is 0.8. The temperature distribution in the lower part of the atmosphere is assumed to be given by the vapor pressure curve. A smooth extrapolation is made to present mean surface temperatures. The change from the high temperature to low-temperature form takes place in a time short compared with geological time. The curve *B* is based upon a time dependence of *G* obtained from the following paper with $\epsilon_0=0.5$. (See Fig. 1 of the following paper.)

Referring to Fig. 1, there is no particular difficulty in accounting for life over a period of the past billion years. Hot springs algae survive at a temperature of 90°C.[9] The high temperature could conceivably be the reason that no life is observed prior to one billion years ago. Also, it has been suggested that the conditions for

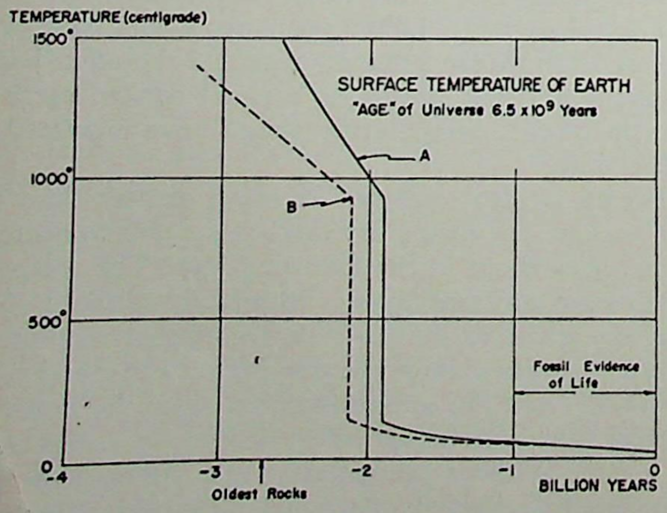


FIG. 1. Temperature curve of earth assuming Dirac's hypothesis.

spontaneous origin of life in a condensed organic goo would be improved at a somewhat elevated temperature of about 100°C.[10]

It is known from radioactive dating that the oldest rocks are not found at a single spot of the earth but are located in such widely separated regions as the Canadian and African shields. These carefully dated rocks all have ages of about 2.7 billion years. This apparent upper limit on the age of terrestrial rocks is much lower than the age of 4.5 billion years found for meteorites. It may be significant that according to Fig. 1, temperatures of the order of the melting point of surface rocks are obtained at about this time.

Finally, there is some geological evidence that the earth was formerly very hot. The separation of the earth into a light crust, heavier mantle and very heavy fluid core lends support to the picture of an originally liquid earth. This is also supported by the fact that the light silicic rocks are found localized as continental masses. The existence of an originally hot sun is also supported by the evidence that the inner planets lost much of their lighter components and that the uncompressed densities of the inner planets range in a regular sequence from 5.2 to 3.8, with Mercury being the heaviest. The hypothesis of an originally hot sun is also supported by the essential absence of the noble gases from the earth's atmosphere. This is usually taken to mean that the earth lost its first atmosphere. Argon is an exception and is believed to have resulted almost wholly from the decay of K_{40}.

FORMATION OF THE MOON

The very interesting problem of the origin of the moon deserves some mention, not because it says anything very conclusive for or against the proposition that the weak interactions were stronger in the past, but because it illustrates the drastic and subtle changes that can be induced in a physical picture by a relatively minor change in an underlying physical law.

The picture of the formation of the earth-moon system which seems to be fairly generally adhered to at present has these two bodies formed apart,[11] and there seems to be a considerable body of opinion which alleges that earth and moon were formed cold.[12] With the assumption that the sun was much hotter in the past, it is much more likely that these bodies were molten at the time of formation and that the moon was derived from the earth by fission.

The diameter of the moon is 0.273 times that of the earth and its mean density is 3.34 against a mean density for the earth of 5.52. It is rigid and its surface departs rather greatly from an equilibrium figure.

Expressed in terms of principal moments of inertia

[9] F. H. Johnson, Sci. American **191**, 65 (1954).

[10] H. F. Blum (private communication).

[11] H. Jeffreys, *The Earth* (Cambridge University Press, Cambridge, England, 1952), third edition.

[12] H. C. Urey, *The Planets* (Yale University Press, New Haven, 1952).

A, B, and C about the axis toward the earth, in the orbit plane and perpendicular to the orbit plane of the moon respectively, the ellipticities are[13]

$$(C-A)/C=(62\pm3)\times10^{-5}$$
$$(B-A)/C=(11.8\pm6)\times10^{-5}.$$

For an equilibrium surface including the effect of tidal forces the ellipticities would be

$$(C-A)/C=3.75\times10^{-5}$$
$$(B-A)/C=2.81\times10^{-5}.$$

The earth is divided into two parts, of which the inner liquid core is believed to be largely iron and roughly half the diameter of the earth. The outer mantle is solid but is moderately plastic and has an uncompressed density of approximately 3.4. It is difficult to explain the low mean density of the moon without assuming that its composition is drastically different from the earth's (uncompressed density $\sim$4.4). The moon would have roughly the right density if it were assumed to have a composition similar to the mantle of the earth. If it were assumed to have been thrown out of the earth in the manner suggested in 1878 by George Darwin, this composition difference would be understandable.

From ancient eclipse observations, it is known that the earth is presently slowing down because of tidal forces and that the moon is receding from the earth to conserve angular momentum. If this be projected into the past it is found that some 3 to 4 billion years ago the moon was very close to the earth (one earth diameter away) and had a period equal to the rotation period of the earth of about 4 hours. It is difficult to carry out this projection with any great confidence as a knowledge of solar-tide angular momentum transfer is required.

Darwin suggested a tidal wave resonance[11] as the mechanism for the fission of a primordial earth. The longest oscillation period of a rigid earth is about 1 hour, but for a liquid earth the longest period would be roughly 2 hours and would be resonant with a 4-hour day on the earth.

If it be assumed that the gravitational constant was twice its present value at the time of the fission of the earth, the period of the liquid earth vibration would be roughly 1.5 hours and the distance of the moon's center from the earth's would be just the diameter of the earth for a period of rotation of 3 hours.

A liquid earth of the present radius radiating as a blackbody would freeze from the bottom up in a few hundred years. On the other hand, it would require at least two orders of magnitude more time to build up a resonant tidal wave in the earth.[11] This much time is required on resonance to produce the requisite energy transfer. Also E. W. Brown[11] has pointed out that the detuning of the resonance produced by nonlinear terms is a source of difficulty. These terms produce a tilted resonance curve which can be climbed only when approached from the right direction. This generally requires a slow variation of the resonance frequency relative to the driving frequency, hence much more time than the minimum discussed above. It would appear that at least 10^5 years would be required to build up the necessary tidal resonance in the earth.

On the other hand, if the temperature of the earth varied in the manner described in the previous section, its temperature would be above the melting point of olivine 3–4 billion years ago. The gravitational energy of formation is easily sufficient to melt the earth at the time of formation. It is clear, therefore, that an abnormally hot sun could have served to keep the earth fluid during the long period required for a tidal resonance to build up.

A possible explanation for the ellipticities of the moon is the assumption that the moon possesses a fossil tidal wave, frozen in when the moon was 70 000 km from the earth.[14] By assuming that the moon was in rotation at the time of solidification and that the rotation was damped out by tidal friction near the end of the solidification process, the observed ellipticities may be accounted for.

Seventy-thousand kilometers is about $\frac{1}{6}$ of the present radius of the moon's orbit, and a very long time would be required for the moon to move that far away from the earth. Radiating as a blackbody, and assuming an adiabatic temperature gradient, the moon would freeze from the bottom up in about 100 years. It is clear that the fossil wave theory of the moon's shape can be made compatible by assuming a hot sun or some other way of keeping the moon warm for a long period of time.

Many years ago Jeffreys showed that turbulent damping would keep a tidal resonance from building up in a primordial earth. However, it has been pointed out by J. A. Wheeler[15] that the earth's magnetic field could suppress turbulence so that only the small viscous damping would be encountered.

The surface of the moon is a most valuable source of information concerning early times. Being devoid of atmosphere there has been little weathering to erase the records of the past. The surface is divided into light and dark colored upland and maria regions.

The uplands are very irregular and contain most of the craters. These craters can be easily ordered in their age. This can be done not only from the way they overlap but also from their form. The very old craters present the appearance of having been formed in a very viscous liquid which flowed after the formation (e.g., the Crater Maginus is old and seems to have flowed).

[13] H. Jeffreys in *The Earth as a Planet*, edited by G. P. Kuiper (University of Chicago Press, Chicago, 1953), p. 49.

[14] R. B. Baldwin, *The Face of the Moon* (University of Chicago Press, Chicago, 1949).

[15] J. A. Wheeler (private communication). The author is grateful for several helpful conversations with Professor Wheeler on the question of the moon's formation.

For the oldest craters there are only vestigial remains in the form of a very shallow circular depression to indicate their presence. There appears to be a correlation between crater size and age with the largest craters being old. This correlation may not be significant, however, as the old small craters could have disappeared from view.

The maria present a very different appearance. Craters here are not nearly so numerous and are usually new in their apearance with little evidence of flow. The maria regions were clearly fluid after the uplands were solid as indicated by the invasion of this dark colored rock into the interiors of some upland craters and the appearance of the boundaries between the two regions.

With the assumption of a hot sun and the present albedo and emissivity, the surface temperature of the moon without an atmosphere would have been near the melting point 3–4 billion years ago. The apparent flow of the old upland craters is believed to be significant. It can be accounted for if the surface were at a temperature only slightly below the melting point of upland rocks for very long periods of time. The surface would be expected to flow under these conditions.

The fact that there are relatively few craters in the maria indicates either a very long interval of time between the formation of the uplands and the solidification of the maria or a very rapid decrease of the rate of meteoric incidence with time. The first possibility would be accounted for with a slowly cooling sun. This would lead to a very slowly cooling moon and a long time interval between the solidification of the uplands and the maria. The second possibility would seem to require a large number of impacts right after the time of formation. It is conceivable that in the fission of the earth-moon system, many small fragments were also formed. These could have been the chief source of the moon's craters.

Because of the rigidity of the moon it would be expected to show numerous small surface cracks produced by the general expansion accompanying a weakening of gravity after the time of solidification. If solidification occurred about 3.25 billion years ago the surface area would be expected to increase about 1% in the intervening period as a result of the halving of the gravitational constant. Silicates are very weak under tension and should produce many small cracks. Many large cracks, called rills, appear in good photographs of the moon.[16]

Cracks appear also in the ocean floors of the earth, and are believed by some geologists to be of relatively recent origin. If the ocean bottoms slowly creep as is claimed by some geologists, it is unlikely that the very slow expansion produced by gravity would produce permanent cracks.

The amount of expansion expected for the earth is very much greater than for the moon. From the known compressibility curves of the mantle and core, one computes an increase in volume of about 15% in the past 3.25 billion years. This corresponds to a change of 4.5% in circumference, or 1100 miles. This would amount to some 50 cracks 20 miles wide distributed over the ocean bottom. If such cracks were permanent, once formed, they would be a prominent feature of the ocean floor. Alternatively, the expansion could have occurred through very small cracks filled with sediment or intrusions, or through "creep" of the ocean bottom.

EARTH'S ROTATION

With a time variation in the gravitational constant there are two ways of defining a time scale. If an atomic measure of time is employed, the moon and planets will be found to be moving more slowly with lapse of time. The ratio of a gravitational to an atomic frequency should vary inversely as the square root of the age of the universe. Atomic clocks are now just being developed that will enable a significant comparison of atomic and gravitational time to be made in an interval of 5–10 years.

Meanwhile, it is possible to make use of an atomic clock provided for us. It is a relatively poor one, but it has been operating for the past 2500 years, and short period irregularities should have somewhat averaged out.

The atomic clock under discussion is the earth and its rotation. The dimensions of the earth are largely determined by nongravitational forces, and its rotation rate should be relatively constant on an atomic time scale. Observations of eclipses of the sun made by Babylonian and Egyptian priests 2500 years ago can be compared with modern observations to give a comparison of atomic and gravitational time.[17]

As was discussed previously, these eclipse measurements show that the earth has been slowing down in its rotation and this change in rotation rate can be traced to a tidal effect. On the other hand, the observations also give the changes in the moon's period. On the assumption that angular momentum lost by the earth in its rotation is transferred to the earth-moon system, it should be possible to predict the motion of the moon.

In order to carry out this calculation it is necessary to know the ratio of solar to lunar induced tidal dissipation. This cannot be properly calculated but was estimated by Jeffreys. The final result is this: the observations are consistent only if it is assumed that the earth is speeding up in an anomalous way at a relative rate of 9×10^{-11}/year, or equivalently that the moon is slowing down relative to the earth.

Assuming a time varying gravity and an age of the Universe of 6.5×10^{9} years, the expected rate of increase of the moon's period is 7.7×10^{-11}/year. This agrees very well with the above observed value.

16 Reference 12, p. 55, also reference 14.

17 H. S. Jones, *The Earth as a Planet*, edited by G. P. Kuiper (University of Chicago Press, Chicago, 1954).

There are, of course, other possible explanations, including a change in the moment of inertia of the earth due to a rain of iron into the core,[12] the melting or growth of ice caps,[17] or the effect of atmospheric tides.[18]

FLOW OF HEAT FROM THE EARTH

It is remarkable that the rate of heat flow from the earth should present a crucial test for a physical hypothesis, but such appears to be the case. This comes about in the following way: It is known that the temperature of the mantle is quite high, believed by many geophysicists to be near the melting point of the mantle. Under these conditions a gradual release of pressure, because of a decreasing gravitational interaction, would lead to a lower melting point. Would the mantle remelt? If it did not, heat would flow from the earth at such a rate as to keep the mantle from melting. If this calculated heat flow greatly exceeded the observed rate of heat flow, the physical hypothesis could be discarded.[19]

Assuming an originally liquid earth, the adiabatic temperature gradient of the upper part of the mantle would be about 1.1°/km (Adams, 1924), and the melting point gradient would be 2.5°–5°/km. Because the melting point gradient exceeds the adiabatic gradient present in the convective mantle, it should freeze from the bottom up.

Uffen (1952) has computed the melting point in the mantle and obtains 4700°C for the bottom of the mantle, assuming 1350°C for 100 km depth.

After the mantle has solidified, the melting point temperature is essentially frozen in. For depths greater than 500 km heat conduction to the surface is of negligible importance.

The outer 100 kms would be cooled by conduction. Its measured temperature gradient is on the average about 20°/km.

Although heat conduction is of negligible importance, convection in the solid mantle is probably of considerable importance in transferring heat to the surface. If convective equilibrium existed in the solid mantle, the temperature gradient would be quite small, leading to a temperature at the core boundary of only 1400°–2500°C depending upon assumptions (Birch, Verhoogen). In view of the observed rigidity of the moon, it is unlikely that the viscosity of the mantle would be sufficiently low for convection to be very important at temperatures this far from the melting point. Also, there would be difficulty in accounting for a liquid iron core (Verhoogen).

After the mantle has solidified, its temperature is at the melting point, and its temperature gradient greatly exceeds the adiabatic convective gradient of

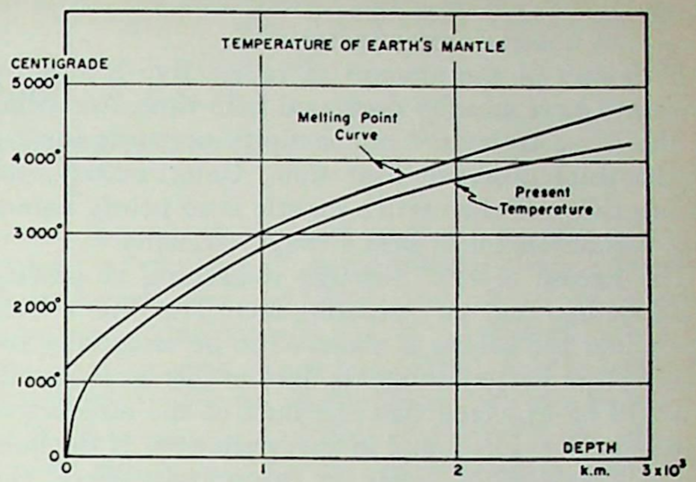


FIG. 2. Temperature distribution in the earth's mantle assuming limiting of convection, by viscosity.

the solid mantle. The mantle would have a relatively low viscosity ($\ll 10^{22}$) at the melting point and convection would occur.

The convective cooling of the mantle would occur until the temperature fell some hundreds of degrees below the melting point after which the viscosity would be too high for convection to continue.

If there were a source of heat in the interior of the earth, the temperature of the mantle would adjust itself at each depth to allow convection to occur at such a rate as to carry the necessary amount of heat away. Hence, the temperature curve below the 100-km level should be such that viscosity would be roughly constant and lie some hundreds of degrees below the melting point curve.

This convection mechanism would occur for depths greater than 100 km, with conduction being the dominant heat transfer mechanism in the first 100 km. An estimated temperature curve is plotted in Fig. 2.

With this heat transfer mechanism in mind, it is easy to discuss the effect of a gradual decrease in the gravitational constant. At any given depth the melting point would decrease with decreasing pressure and this would result in a decrease in the viscosity of the mantle. This would result in increased convection, with a transfer of heat to the surface to keep the viscosity at the proper value.

A knowledge of the melting point curve, pressure curve, and the specific heat of the mantle and core is sufficient to enable a calculation of the rate of heat flow from the earth. Assuming that the gravitational constant, hence pressure, varies inversely as the age of the Universe (6.5×10^9 years) and taking the average specific heat of the mantle and core to be 0.2 and 0.15, respectively, this contribution to the rate of flow of heat from the earth is found by numerical integration to be 1×10^{-6} cal/cm² sec. This agrees with the total of 1×10^{-6} cal/cm² sec actually observed.

In view of the uncertainty regarding the temperature curve and the specific heats, the effect of the variation

[18] Holmberg, Monthly Notices Geo. Suppl. 6, 325 (1952).

[19] The references for the following discussion are to be found in an excellent survey of the literature by J. Verhoozen, *Physics and Chemistry of the Earth*, Ahrens *et al.*, editors (McGraw-Hill Book Company, Inc., New York, 1956).

of gravity in releasing the earth's heat should be regarded as uncertain by a factor of 2.

Estimates of the amount of radioactive heating of the earth have steadily decreased with time. According to the latest estimate[20] radioactivity accounts for 2/3 of the total observed heat flow. Unfortunately, the radioactivity of the earth's mantle is so poorly known that this number is at best a rough estimate.

The release of heat through weakening of gravity could explain one very puzzling fact. The rate of heat flow under the oceans is observed to be essentially the same as that in the continents. But on chemical grounds it would be expected that the bulk of the radioactive materials would be found in the continents. If the bulk of the heat had its origin in the gravity effect, the uniformity of distribution is obvious.

The gravitational release of heat by the mantle is accompanied by a gradual lowering of the temperature of the core. This contribution to the outward heat flow was included in the foregoing number.

Because of the varying gravitational constant, heat flows out of the core itself at a rate of about 10^{12} cal/sec. In order to maintain sufficient convection in the core to account for the earth's magnetism, a heat flow out of the core of at least 3×10^{11} cal/sec[21] is required. That there should be enough radioactivity in the core to account for such heat release is unlikely in view of the great chemical activity of the radioactive materials.

To summarize, it appears that the heat flow from the earth that would accompany a varying gravitational constant is not excessive and furthermore could be the chief source of the observed flow. This could account for the heat flow under the oceans. It could also account for the convection in the earth's core which is needed to explain the earth's magnetism.

[20] H. C. Urey, Proc. Natl. Acad. Sci. 42, 889 (1956).
[21] W. Elsasser, Revs. Modern Phys. 22, 29 (1950).

BETA DECAY AND GEOLOGICAL DATING

The second column of Fig. 1 contains the square of the reciprocal of the Fermi interaction constants. Following Dirac's rule this would imply that β decay rates would vary inversely as the square root of the age of the Universe. For 2 billion year-old rocks there should be a 10% discrepancy between rubidium-strontium dates and those obtained from uranium or thorium. The strontium dating method should give ages 10% greater than those obtained from α decay measurements.

It is doubtful that present geological dating methods are sufficiently reliable to detect an absolute difference of 5–10% between two quite different dating methods.[22]

To summarize, none of the evidence reviewed here can be used to give strong support to Dirac's hypothesis. On the other hand, it would appear that a variable gravitational interaction cannot be easily excluded by evidence reviewed here.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge helpful conversations with H. Hess of the Geology Department, F. Johnson of the Biology Department, and L. Spitzer and M. Schwarzschild of the Astronomy Department, all of Princeton University.

[22] See L. H. Ahrens, *Physics and Chemistry of the Earth*, for a recent survey article on this subject.

Gravitation without a Principle of Equivalence

R. H. Dicke

Palmer Physical Laboratory, Princeton, New Jersey

THE previous article has considered the observational and experimental facts and has concluded that there is no substantial evidence to support the belief that the coupling constants of the weak interactions are independent of time or place. Consequently, it is possible that the principle of equivalence may be satisfied, if at all, only when the contributions to the binding energy of a system having their origin in the weak interactions are neglected. This article considers a form which a theory of gravitation may take when the principle of equivalence is satisfied in a weakened form only.

Jordan has previously considered a similar problem,* and Fierz has made a critical analysis of Jordan's theory.†

The great difficulty with constructing a theory of gravitation is the paucity of experimental evidence. After 40 years there are still only the four famous observational checks of the theory of relativity. Of these only two have any real accuracy. With so few experimental facts to guide one, any number of *ad hoc* theories can be constructed. To choose between them, standards going beyond the observational evidence must be introduced. The danger of judging a theory on the basis of elegance, simplicity, or perfection is obvious.

While "elegance" may not be a valid criterion for judging a theory, there are a few rules for the construction of a formalism which if followed should improve the prospects for later agreement with observation. First, it should be noted that there is much experimental evidence on the validity of the Lorentz invariance of the strong interactions and a theory of gravity should reduce to the usual Lorentz invariant theory in the absence of the gravitational field. Second, the theory should introduce a minimum of new elements. Third, on the assumption that nature is basically simple, the simplest of several alternatives should be chosen. The theory to be described accepts Mach's principle, the cosmological principle, and is generally covariant. Also as much of the principle of equivalence as is supported by the Eötvös experiment is accepted.

The general features of a theory of gravitation without a principle of equivalence are easily outlined. The motivation for introducing a Riemannian metric into the geometry of space and time is now largely absent, as there is no single universal gravitational acceleration at a given space-time point. Simply by redefining units of length and time as functions of space-time coordinates, the curvatures of a space are modified. With a proper redefinition of units making them dependent upon coordinates and orientation of an infinitesimal interval a curved space can be converted into a flat one and vice versa. Rosen[1] has shown how to formulate general relativity within the framework of a flat metric.

To illustrate the arbitrary character of the choice of metric tensor, consider the following physical example. Twelve identical rods can be normally assembled into a hexagonal pattern with 6 rods joining at the center. If this assembly is carried out in a suitable gravitational field, the 6 rods no longer join at the center. There are at least two geometrical explanations for this result. The conventional one is that the rods have not changed but are now in a curved space which "causes" a gap to open at the center of the geometrical figure. Another possible explanation is that the gravitational field has shortened the radial rods relative to the circumferential ones. The two explanations are equivalent in the sense that they both agree as to the existence of the gap in the geometrical figure constructed out of real atoms.

It has been argued[2] that space is "really" curved and that the rods do not change their "real" length. Without splitting hairs over the meaning of the word "really" this argument is based on the assumption that such changes are presumably independent of the material out of which the rods are constructed. However, all rods are constructed from electrons, protons, and neutrons held together almost completely by the strong interactions. They have a common structure and could vary in length in a common way.

Two theories which differ only in the definition of the units employed are equivalent. Nonetheless, there are advantages to be obtained from the use of a flat metric. In the conventional theory the Riemannian metric is used to transform the gravitational force away. In a flat metric the gravitational effects are to be regarded as associated with a force field just as electromagnetic or meson forces are related to a field. Because of this common basis it becomes possible to make use of analogy in constructing a theory of gravity.

By analogy with electromagnetic or meson force fields it is reasonable to expect that when viewed sufficiently closely the gravitational effects would be quantum in nature. Stated more exactly, it might be expected that the gravitational force acting on an elementary particle would have its origin in a local interaction of the particle with virtual particles present in the vacuum.

* P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1955), second edition.
† M. Fierz, Helv. Phys. Acta **29**, 128 (1956).

[1] N. Rosen, Phys. Rev. **57**, 147 (1940).
[2] H. P. Robertson, *Albert Einstein*, edited by Paul A. Schilpp (Tudor, New York, 1951).

Remembering the virtual electron-positron pairs reputed to be present in the vacuum as a result of zero point fluctuations, it becomes interesting to inquire whether the gravitational effect can be linked to these particles already present. If so, it should be possible eventually to construct a theory of particles and obtain the gravitational interaction as a weak effect connected with more primitive strong interactions. A less ambitious approach would be to start in the middle of the problem, to ignore the quantum aspects of the interaction of a particle with a bath of virtual particles, and to treat this interaction as a classical field.

The most striking effect of the presence of virtual pairs in the vacuum is the polarizability of the vacuum. This property suffers from divergence difficulties which are usually ameliorated by "renormalization." By defining the velocity of light in empty space as c and "renormalizing," the vacuum polarization effects are made to disappear for a weak electromagnetic wave in free space whereas they still contribute to the space charge about a charged particle. This, however, is arbitrary. The velocity of light in a "bare" space could be greatly different from c or even meaningless.

With a "cutoff" theory there are no divergences and the vacuum polarization can be treated as physically meaningful. It may be significant that a wavelength cut off at the gravitational radius of a particle would not be detectable in any presently known experiment. It should be noted that with such a cutoff theory, the velocity of light in "bare" space is not c. It is c only after including vacuum polarization effects.

It is helpful to remember that before Lorentz no distinction was made between the polarization of a vacuum and the polarization of a dielectric medium. Lorentz first clearly saw that the $\dot{D}$ term in Maxwell's first equation should, on physical grounds, be separated into two parts with the $4\pi\dot{P}$ term representing polarization currents in the dielectric medium. These polarization currents were to be treated like all other currents in computing the magnetic field. The $\dot{E}$ term, on the other hand, was commonly thought to be physically quite different. In the early 20th Century, the vacuum by definition contained no charges and currents. In recent years vacuum polarization effects have been recognized as existing and having physical importance. It would appear to be reasonable to assume that the $\dot{E}$ term represents the flow of virtual charged particles in the vacuum and that the whole of $\dot{D}$ is to be interpreted as a "displacement current."

With the neglect of quantum effects the polarizability of the vacuum can be described by classical field quantities ϵ and μ.

From this point of view the vacuum is to be treated like a dielectric medium. In free space far from particles, it is a medium without dispersion as there is no mechanism which would lead to the absorption of a single photon and the production of a pair.

If the vacuum is thus considered to have a structure, to be a dielectric medium which can be polarized by electromagnetic fields, it becomes important to inquire about the effect of motion relative to such a medium. Could such a motion be measurable by electromagnetic means? The motion of an ordinary dielectric medium leads to measurable effects (Fizeau). That such motional effects should be absent for the vacuum is reasonable on the basis of the following physical argument.

Consider an electron moving through a normal dielectric medium such as helium gas. The electric field about the electron can be expanded as a Fourier integral in time. Up to a frequency of about 10^{15} cycles/sec the helium is equivalent to a moderately continuous medium of constant dielectric constant ϵ. There is a Lorentz contraction in this part of the electric field by a Lorentz factor computed by taking the velocity of light to be the wave velocity in the medium. On the other hand, the parts of the field associated with frequencies which are very much higher (above the strong absorption frequencies) are not greatly affected by the helium gas and the field is contracted by a factor computed for the velocity in free space. The energy stored in the low-frequency part of the electromagnetic field contributes in an anomalous way to the effective mass of the electron but this contribution is small compared with the total mass.

Consider the situation when there is no dispersion in the medium. All frequency components are contracted by a Lorentz factor computed with the velocity of light equal to the wave velocity in the medium. It would be reasonable to expect that the theory would be Lorentz covariant using the wave velocity in the medium as the velocity of light. It should not be possible to use an electromagnetic effect to determine a velocity relative to such a nondispersive medium.

GRAVITATION AN ELECTROMAGNETIC EFFECT?

The fact that many of the properties of gravitation can be accounted for in terms of an interaction with a polarizable medium is an old idea which has recurred from time to time.[3] The physical idea is simply that a space variation in the polarizabilities of the vacuum will lead to a number of results familiar as typical gravitation effects. For example, an increase of the index of refraction of the vacuum in the vicinity of the sun will cause a bending of light toward the sun.

The gravitational force on a charged particle is interpreted as resulting from a change in its electromagnetic self-energy with position as a result of a variation in the polarizability of the vacuum. A gradient in the polarizability results in a force acting on the charged particle. This force results in part from the polarization charges induced in the vacuum by the charged particle. For a medium having a nonvanishing

[3] H. A. Wilson, Phys. Rev. 17, 54 (1921) introduced many of the physical concepts employed in this paper.

gradient in its polarizabilities there is more induced charge on one side of the particle than on the other and the electrostatic interaction with the induced charges leads to a force acting on the particle in the direction of increasing gradient.

For a complex system such as a hydrogen atom there is an added force having its origin in a variation of the binding energy of the atom with position. The neutron is to be regarded as a compound system containing charged particles.

As a result of the change of binding energy with position it would be expected that the energy or frequency of a photon emitted by the atom would depend upon the location of the atom in the field. This could conceivably lead to the phenomenon of the gravitational red shift.

Also, it would be expected that as a result of the variation in the dielectric constant of the vacuum the Bohr radius and other atomic lengths would be a function of position. This would lead to a shortening and bending of meter sticks depending upon their location. If such meter sticks are defined as unchanged, the resulting metric of space-like surfaces is Riemannian.

Thus there is a possibility of accounting for all the observed gravitational effects within the physical framework outlined above. That gravitation should be electromagnetic in origin is not unreasonable. In the previous paper it was shown that the strength of the gravitational interaction appears to be related to the strength of the electrical interaction and to the size of the universe.

In order to find a functional relation between the dielectric constant ϵ and μ, the permeability of the vacuum, we consider quantitatively the effect of a polarizable medium on the hydrogen atom. The binding energy can be written as

$$\tfrac{1}{2}M_e\alpha^2, \tag{1}$$

where M_e is the electron's rest energy and

$$\alpha=\frac{e^2}{\epsilon\hbar c} \tag{2}$$

is the fine structure constant. e is the "true" charge of the electron and assuming the validity of Maxwell's equations (charge conservation), it is a constant. The dielectric constant ϵ is present in (2) because of the effect of the dielectric medium in reducing the strength of the electrostatic interaction between the electron and proton. $\hbar$ is also a constant if angular momentum is to be conserved. This can be seen by considering a circularly polarized photon carrying an angular momentum $\hbar$. As it propagates through space $\hbar$ remains unchanged if angular momentum is to be conserved.

The velocity of light in a polarizable medium can be written with a suitable choice of units as

$$c=(\epsilon\mu)^{-\frac{1}{2}}, \tag{3}$$

where μ is the permeability of the medium. In the previous paper it was shown that there are experimental and observational reasons for believing that the fine structure constant (2) varies, if at all, only slowly with position and time. Hence, we shall assume that α is a constant. In view of the fact that we are attempting to associate the gravitational field with a gradient in ϵ, from (2) and (3), α can be constant only if

$$\epsilon=\mu. \tag{4}$$

The variation in index of refraction about the sun that is required to obtain a deflection of light of the amount expected and observed (roughly) may be calculated using Maxwell's equations. The result is

$$\epsilon=\mu\cong 1+\frac{2GM}{r}, \tag{5}$$

where G is the gravitational constant based on an energy measure of mass, M is the sun's mass in energy units, and r is the distance from the sun. The second term on the right-hand side of (5) is clearly associated with the presence of the sun.

What about the first term? Does it have its origin in the remainder of the matter in the universe? To investigate this possibility we form the integral

$$2G\int_0^R \frac{4\pi\rho r^2}{r}dr=4\pi G\rho R^2, \tag{6}$$

where ρ is the density of matter in the universe and R is the Hubble radius. Inserting the appropriate value for G and assuming $R=5.4\times10^{27}$ cm, $\rho/c^2=4\times10^{-29}$ g/cm^3, (6) is equal to unity. This density of matter is compatible with astronomical observations which give a value of $\rho/c^2=10^{-30}$ g/cm^3 for galactic matter only.

From the standpoint of Mach's principle this is a highly satisfactory result. In this interpretation, the polarizability of the vacuum at any point depends upon the distribution of the distant galactic matter. As the inertial properties of matter are associated with these polarizabilities, the inertial properties are, as suggested by Mach, determined by the remainder of the matter of the universe.

COORDINATES AND UNITS

Of necessity the preceding discussion has been somewhat inexact, with the physical quantities not carefully defined. By basing the discussion upon a flat space-time, the question of how physical units are to be defined has been left somewhat vague. Atomic units of length and time are no longer suitable local standards as these lead to a Riemannian metric. While it is conceivable that units of length and time might be introduced simply by requiring that the resulting space be flat, the resulting definition would be generally not unique. For example, a stereographic projection may

be used to map a Cartesian coordinate system of a plane upon a sphere and with a suitable definition of a local unit of length the metric of the sphere (with the singular point excluded) is flat, but the choice of the singular point on the sphere is arbitrary and the resulting definition of the unit of length is not unique.

What is needed is a prescription for correcting local atomic units of length, time, etc., so that they give a "true" measure. Most of the problem is already solved. The assumptions that e and $\hbar$ are constant means that these two quantities can serve as "true" local units. However, a third unit is needed. In view of the assumption that ϵ affects atomic lengths and times, what is needed is a local measure of ϵ. It will appear as a derived result of the theory that ϵ is in principle locally measurable by determining locally the ratio of electrical to gravitational forces. Thus, a flat metric is obtained if local units of length and time are assumed to be given by local atomic units corrected by a function of ϵ.

The system of units used with a flat metric will be called "Newtonian." Local atomic units will be called "atomic" or sometimes "proper." As explained later a characteristic atomic length expressed in Newtonian units varies with ϵ as

$$L=L_0/\epsilon^{\frac{1}{2}}, \tag{7}$$

also a characteristic atomic frequency varies as

$$\omega=\omega_0/\epsilon^{\frac{3}{2}}, \tag{8}$$

with L_0 and ω_0 constants.

In addition to the problem of defining local measures of physical quantities it is necessary to discuss coordinate systems. In view of the flat space-time it is always possible to choose Cartesian systems of coordinates. Furthermore with the assumption of the Cosmological principle it is possible to define for every point a unique time direction (cosmic time). It is assumed that this time direction is everywhere parallel to the time axis of a particular Cartesian coordinate system. This Cartesian coordinate system, which for obvious reasons will be called "Newtonian," can be characterized by saying that along the world line of any fixed position point the universe appears uniform.

Because of this high degree of symmetry, the vacuum should appear isotropic in this coordinate system and the vacuum polarizabilities should be scalars. Local inhomogeneities would destroy this symmetry and it is conceivable that the polarizability about a body such as the sun would be a tensor quantity. On the other hand, the simplest assumption to make is that the polarizability is a scalar even in the presence of local irregularities in the mass distribution, and this assumption will be made here.

Although a special coordinate system has been introduced, it has been done within the framework of Mach's principle. Namely this special coordinate system is determined by the distribution of matter in the universe. Furthermore, the equations of motion can be written in a generally covariant manner. The equations do, however, take on a particularly simple physically understandable form for the Newtonian coordinate system.

Units are so chosen that the metric tensor of the Newtonian coordinate system with Newtonian units is γ_{ij}

$$\gamma_{11}=\gamma_{22}=\gamma_{33}=-1,\quad \gamma_{44}=1,\quad \gamma_{ij}=0,\quad i\neq j. \tag{9}$$

Although the physical significance of the equations is more transparent when expressed in Newtonian units it is always possible to choose proper or local atomic units if desired. In this case assuming for the moment the validity of (7) and (8) the metric tensor for this particular coordinate frame is g_{ij}

$$g_{11}=g_{22}=g_{33}=-\epsilon,\quad g_{44}=1/\epsilon,\quad g_{ij}=0,\quad i\neq j. \tag{10}$$

Whether g_{ij} is taken to be an ordinary tensor or the metric tensor, the infinitesimal invariant

$$ds^2=g_{ij}dx^idx^j, \tag{11}$$

is a useful quantity representing an infinitesimal interval measured in local atomic units.

POSTULATE OF LOCAL LORENTZ COVARIANCE

Before proceeding further we shall summarize the assumptions which are being made.

1. The theory is based upon a flat metric.
2. Mach's principle is assumed.
3. The Cosmological principle is assumed.
4. A special Cartesian coordinate system ("Newtonian") determined by the distribution of matter in the universe is introduced such that the universe as a whole appears isotropic from any fixed position point. While generally covariant equations can and will be written, the Newtonian coordinate system because of the cosmological principle has a special physical significance.
5. The vacuum is polarizable and Maxwell's equations for a polarizable medium are valid in the Newtonian coordinate system. The electric and magnetic specific inductive capacities of the vacuum ϵ, μ are scalars.
6. A variation in ϵ, μ affect the self-energy of a particle. Characteristic atomic lengths, frequencies and energies are affected by a variation in ϵ and μ.
7. e and $\hbar$ are constant. (Conservation of charge and angular momentum.)
8. Neglecting the small contribution to the binding energy of a complex system having an origin in the weak interactions, the self-energy of a particle, whether elementary or complex, varies with ϵ as

$$M=M_0\epsilon^{-\frac{1}{2}},\quad M_0=\text{constant}.$$

It is shown later that assumption 8 can be obtained from a limited principle of equivalence. Also the fact that all atoms obey the same functional relation is

justified directly by the Eötvös experiment (see the preceding paper).

From the above assumptions several important results are obtained.

9. The fine structure constant

$$\alpha=\frac{e^2}{\epsilon\hbar c}$$

is constant (independent of ϵ and μ). Also all other development parameters of strong interactions are constant. This follows if both the self-energies of the elementary particles and complex atoms are to obey No. 8.

10. From 9 it follows that $\epsilon=\mu$.

11. From 8, 9, and 10 it follows that all characteristic atomic lengths vary together as

$$L=L_0\epsilon^{-\frac{1}{2}},\quad L_0=\text{constant}.$$

[For example compute the Bohr radius as

$$a_0=\frac{e^2}{\epsilon M_e}\alpha^{-2}.]$$

12. From 8 and 7 all atomic frequencies vary as

$$\omega=\omega_0\epsilon^{-\frac{1}{2}}$$

(since $\hbar$ is constant and $M\sim\hbar\omega$).

The above list of assumptions by itself would constitute a weak scaffolding upon which to construct a theory of gravity. Presumably a large number of *ad hoc* theories could be constructed within this framework to account for the very few observational facts. However, there is a strong assumption still to be made which greatly limits the number of possibilities.

Physical arguments were given earlier to support the assumption that an atom moving relative to the Newtonian frame would experience a time dilatation and Lorentz contraction in accordance with the local velocity of light. This suggests that equations be written in a way which we define as local-Lorentz covariant. Stated exactly the Lagrangian or Lagrangian density of a field theory should be written in its usual Lorentz invariant form but with c no longer constant but equal to the reciprocal of ϵ.

Since this Lagrangian density reduces for constant ϵ to the usual Lorentz invariant density, all the observational experience concerning special relativity can be brought to bear on the problem of choosing the correct Lagrangian density.

EQUATIONS OF MOTION

Rather than consider immediately the formulation of the general equations we first limit ourselves to a simple problem: Find the motion of a test particle in a given gravitational field. The scalar ϵ is assumed known and given. Following the prescription outlined above the Lagrangian is written for the Newtonian coordinate system as a local Lorentz invariant.

$$L=-M(1-\epsilon^2v^2)^{\frac{1}{2}}=-M_0\epsilon^{-\frac{1}{2}}(1-\epsilon^2v^2)^{\frac{1}{2}}. \tag{12}$$

Remember that the velocity of light $c=\epsilon^{-1}$. M is the rest energy of the particle. The equations of motion are obtained from the variational principle

$$\delta\int_{t_1}^{t_2} Ldt=0,\quad dt=dx^4. \tag{13}$$

This gives for Euler equations

$$\frac{d}{dt}\left[\frac{M\epsilon^2\mathbf{v}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\right]=\tfrac{1}{2}M\left[\frac{1+\epsilon^2v^2}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\right]\frac{1}{\epsilon}\nabla\epsilon. \tag{14}$$

The momentum of the particle is

$$\mathbf{p}=\frac{\partial L}{\partial\mathbf{v}}=\frac{M\epsilon^2\mathbf{v}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}, \tag{15}$$

and the Hamiltonian is

$$H=\mathbf{p}\cdot\mathbf{v}-L=\frac{M}{(1-\epsilon^2v^2)^{\frac{1}{2}}}. \tag{16}$$

It is a constant of the motion, (the energy of the particle) if ϵ is time independent. For a time dependent ϵ the rate of change of the energy of the particle is

$$\frac{dH}{dt}=-\frac{\partial L}{\partial t}=-\frac{1}{2}\frac{M}{(1-\epsilon^2v^2)^{\frac{1}{2}}}(1+\epsilon^2v^2)\frac{1}{\epsilon}\frac{\partial\epsilon}{\partial t}. \tag{16a}$$

Note that from Eq. (23) the gravitational force $\mathbf{F}$ is

$$\mathbf{F}=\frac{1}{2}\frac{M}{(1-\epsilon^2v^2)^{\frac{1}{2}}}(1+\epsilon^2v^2)\frac{1}{\epsilon}\nabla\epsilon. \tag{17}$$

The force acting on the particle is proportional to the particle's energy (ϵ time independent), but it contains the added factor $(1+\epsilon^2v^2)$. This added factor serves to double the gravitational force for a rapidly moving particle. This leads to double the transverse gravitational acceleration for a photon and twice the Newtonian deflection of light by the sun. This result is in agreement with observations.

Although a rapidly moving particle has double the normal weight, this does not contribute anomalously to the weight of a bound system. Consider a photon confined to a box with perfectly reflecting walls. The photon is too heavy by a factor of 2. However, a small displacement of the box upward results in a change in volume of the box because of expansion of the walls. The photon gas does work on the walls as they expand and the resulting work done by the photon is just sufficient to reduce its effective weight to the normal value.

Equation (12) can be written in generally covariant form by introducing the tensor g_{ij} (not the metric tensor) defined by (10) and the infinitesimal invariant of (11). In this notation (12) becomes

$$L=-M_0\frac{ds}{dt}. \quad (18)$$

The variational principle takes on the invariant form

$$\delta\int_{s_1}^{s_2} ds=0, \quad (19)$$

since M_0 is a constant.

Equation (19) is identical with the corresponding equation from the Einstein formalism where the orbit appears as a geodesic. Thus, the only differences between the two theories concern the form of the tensor g_{ij} and not the equations of motion with a given g_{ij}.

HOMOGENEOUS FIELD EQUATION

From the physical picture which has been constructed it is reasonable to expect that the scalar field ϵ should exhibit retardation effects with gravitational waves traveling with the local velocity of light. Consequently, it is assumed that an invariant Lagrangian density for the gravitational field alone can be constructed by writing the usual Lagrangian density for a scalar field, an invariant quadratic in gradients of ϵ

$$L=\frac{1}{2k}h^{ij}\epsilon_{,i}\epsilon_{,j}. \quad (20)$$

Here h^{ij} is the reciprocal of the tensor g_{ij}, i.e.,

$$g_{ik}h^{kj}=\delta_i^j, \quad (20a)$$

and

$$\epsilon_{,i}=\frac{\partial\epsilon}{\partial x^i}, \quad (21)$$

k is a constant. This is the same form as the standard Lorentz invariant Lagrangian density of a zero mass scalar field and appears to be the simplest invariant Lagrangian density which will yield gravitational waves of the appropriate velocity. While (20) could be multiplied by an arbitrary function of the scalar ϵ without destroying its invariance, this is not done, and the physical reason for omitting such a factor is discussed later.

The gravitational wave equation for matter free space is obtained from the variational principle

$$0=\delta\int L\sqrt{-\gamma}d^4x. \quad (22)$$

Here as usual $\sqrt{-\gamma}d^4x$ is the invariant volume element. The wave equation is obtained as the Euler equation

$$\frac{\partial}{\partial x^i}\left((-\gamma)^{\frac{1}{2}}\frac{\partial L}{\partial\epsilon_{,i}}\right)-\sqrt{-\gamma}\frac{\partial L}{\partial\epsilon}=0. \quad (23)$$

Conservation laws for the gravitational field alone are obtained by defining the energy momentum tensor $G_j{}^i$ as

$$G_j{}^i=\epsilon_{,j}\frac{\partial L}{\partial\epsilon_{,i}}-\delta_j{}^iL. \quad (24)$$

The covariant divergence (γ is the metric tensor) of the energy momentum tensor vanishes. The proof of this is simple. For the Newtonian-Cartesian coordinate system the Christoffel symbols all vanish and the covariant divergence and ordinary divergence are equal

$$G_{j,\,i}{}^i=G_{j;\,i}{}^i. \quad (25)$$

However, for this coordinate system $\sqrt{-\gamma}=1$ and making use of (23) it is found that

$$G_{j,\,i}{}^i=0. \quad (26)$$

The vanishing of the covariant divergence for this coordinate system implies its vanishing in all coordinate systems since

$$G_{j;\,i}{}^i=0 \quad (27)$$

is a tensor equation.

The vanishing of the ordinary divergence (26) for the Newtonian coordinate system can be interpreted in the usual way as representing the conservation of energy and momentum of the field. Equation (27) represents the same conservation laws expressed in general curvilinear coordinates.

In the Einstein theory the vanishing of the covariant divergence of an energy-momentum tensor does not represent a conservation relation as (26) does not exist in this formalism. In the present theory, however, the existence of at least one coordinate system for which the covariant divergence equals the ordinary divergence and vanishes is sufficient to define conservation relations which can be expressed in any arbitrary coordinate system.

The components of the tensor $G_j{}^k$ are obtained from (24) as

$$G_j{}^i=\frac{1}{k}h^{ik}\epsilon_{,k}\epsilon_{,j}-\delta_j{}^i\frac{1}{2k}h^{kl}\epsilon_{,k}\epsilon_{,l}. \quad (27a)$$

Written out for the Newtonian coordinate system the energy density (positive definite) is

$$G_4{}^4=\frac{1}{2k}\left[\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)^2+\frac{1}{\epsilon}(\nabla\epsilon)^2\right]. \quad (27b)$$

The three components of momentum density are

$$-G_\alpha{}^4=-\frac{1}{k}\epsilon\epsilon_{,4}\epsilon_{,\alpha}=-\frac{1}{k}\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)\left(\frac{\partial\epsilon}{\partial x^\alpha}\right), \quad (27c)$$

$\alpha=1, 2, 3.$

The three components of the energy flux density are

$$G_4{}^{\alpha}=-\frac{1}{k\epsilon}\left(\frac{\partial\epsilon}{\partial t}\right)\left(\frac{\partial\epsilon}{\partial x^{\alpha}}\right)=-\frac{1}{\epsilon^2}G_{\alpha}{}^4. \quad (27d)$$

The remaining elements of $G_j{}^i$ constitute the stress tensor of the field.

THE INHOMOGENEOUS FIELD EQUATION

By adding to the Lagrangian density (20) terms of the type of (12) a Lagrangian density representing particles, the gravitational field and their interaction is obtained. This may be written for the Newtonian frame as

$$L=-\sum_i M_i(1-\epsilon^2 v_i{}^2)^{\frac{1}{2}}\delta(\mathbf{r}-\mathbf{r}_i)+\frac{1}{2k}h^{ij}\epsilon_{,j}\epsilon_{,j}. \quad (28)$$

Here $\mathbf{r}$ represents the three position components of the field point. $\mathbf{r}_i$ refers to the position of the ith particle and $\mathbf{v}_i$ to its velocity.

The variational principle (22) gives as the Euler equation for ϵ

$$\frac{\partial}{\partial x^i}\frac{\partial L}{\partial\epsilon_{,i}}-\frac{\partial L}{\partial\epsilon}=0, \quad (29)$$

which becomes, when written out in detail for the Newtonian coordinate system,

$$\nabla^2\epsilon-\epsilon^2\frac{\partial^2\epsilon}{\partial t^2}=\frac{1}{2}\left[\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)^2+\frac{1}{\epsilon}(\nabla\epsilon)^2\right]-\frac{k}{2}\sum_i\frac{M_i(1+\epsilon^2 v_i{}^2)}{(1-\epsilon^2 v_i{}^2)^{\frac{1}{2}}}\delta(\mathbf{r}-\mathbf{r}_i). \quad (30)$$

The first term on the right is easily seen from (27b) to be $kG_4{}^4$ and represents k times the gravitational energy density. The material particles and gravitational energy serve as sources of gravitational waves.

That a particle should be a source of a gradient of ϵ is understandable when it is recalled that a gradient of ϵ leads to a force acting on the particle. It would be expected on grounds of Newton's third law that the particle would be a source of the field which acts upon it.

MECHANICS OF CONTINUOUS MATTER

The dynamics of point particles moving under the influence of gravitational forces as described by (14) and (30) is not completely satisfactory. It is necessary to subtract the self-gravitational field of a particle before computing the force on a particle or the total gravitational energy of the field. If this is not done, divergent results are obtained. The difficulty is of essentially the same type as the self-field problem encountered in the Lorentz-Maxwell theory. There is one important difference, however; the gravitational field (30) is nonlinear and the subtraction cannot be carried out in a consistent fashion. On the other hand the nonlinearities become important only at a distance from a particle of the order of the gravitational radius.

From a formal viewpoint matter is best considered to be a continuous medium, as the self-field problem then disappears. This of course is nothing but a way of ignoring the particle-structure problem which cannot be solved in any case.

The field equations for a continuous medium moving under the influence of gravity only are conveniently written in terms of energy momentum tensors. Inasmuch as the equations of motion of a particle in a given gravitational field (g_{ij}) are identical with the equations of motion of the Einstein theory, these equations can be obtained from the latter theory in the form of the vanishing of the covariant divergence of the energy momentum tensor of matter with zero stresses. Owing to the difference in the units employed with the metric tensors γ_{ij} and g_{ij}, the matter tensor of the Einstein theory must, for the Newtonian frame, be multiplied by $\sqrt{-g}$ to obtain the energy momentum tensor for matter in the electromagnetic theory $M_i{}^k$. Hence, the Einstein equations of motion are

$$\left(\frac{1}{\sqrt{-g}}M_i{}^k\right)_{;k}=0, \quad (31)$$

where the covariant divergence is computed with g_{ij} as the metric tensor.

$$M_i{}^k=\rho g_{ij}u^j u^k, \quad (32)$$

where ρ is the matter energy density in a coordinate frame for which the matter is locally at rest.[4] u^i is the four velocity dx^i/ds. Making use of a standard formula the covariant divergence (31) can be written as

$$0=\sqrt{-g}\left(\frac{1}{\sqrt{-g}}M_i{}^k\right)_{;k}=M_{i,k}{}^k-\frac{1}{2}\frac{\partial g_{lm}}{\partial x^i}h^{ln}M_n{}^m. \quad (33)$$

Making use of (10), (11), and (20a) the last term in (42) can be evaluated for the Newtonian frame to give

$$-\frac{1}{2}\frac{\partial g_{lm}}{\partial x^i}h^{ln}M_n{}^m=-\frac{1}{2}\frac{\rho}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{1+\epsilon^2v^2}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{1}{\epsilon}\epsilon_{,i}. \quad (34)$$

Because of the Lorentz contraction the matter density is increased by a Lorentz contraction factor. Comparing (34) with (17) it is apparent that (34) represents the gravitational force density for $i=1, 2, 3$. For $i=4$ it represents the negative of the rate per unit volume that work is being done on matter by gravitational forces [see (16a)].

The first term on the right of (33) represents, for the Newtonian frame, the total derivative (i.e., co-moving)

[4] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Press, Cambridge, 1951), p. 296.

with respect to time of the density of matter momentum and energy. To see this (32) is written as

$$M_i{}^k=\rho g_{ij}\frac{\epsilon v_j v_k}{(1-\epsilon^2v^2)}, \tag{35}$$

with v_α representing an ordinary velocity component $\alpha=1, 2, 3$ and $v_4=1$

$$-M_{\alpha,k}{}^k=\frac{\rho\sqrt{\epsilon}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{d}{dt}\left[\frac{\epsilon^{\frac{3}{2}}v_\alpha}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\right]$$

$$+\frac{\epsilon^{\frac{3}{2}}v_\alpha}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\left[\frac{\rho\sqrt{\epsilon}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}v_k\right]_{,k}. \tag{36}$$

Also the condition of matter continuity is

$$\left[\frac{\rho\sqrt{\epsilon}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}v_k\right]_{,k}=0. \tag{36a}$$

Equation (36a) can be understood from a particle model for the continuous medium. The coefficient of v_k is proportional to the particle number density and this divergence vanishes if numbers of particles are conserved. Hence

$$-M_{\alpha,k}{}^k=\frac{\rho\sqrt{\epsilon}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{d}{dt}\left[\frac{\epsilon^{\frac{3}{2}}v_\alpha}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\right], \tag{37}$$

$$\alpha=1, 2, 3.$$

In similar fashion

$$M_{4,k}{}^k=\frac{\rho\sqrt{\epsilon}}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{d}{dt}\left[\frac{1}{\sqrt{\epsilon}(1-\epsilon^2v^2)^{\frac{1}{2}}}\right]. \tag{37a}$$

These equations represent total time derivatives of the matter densities of momentum and energy respectively.

Comparing (33), (34), (37), and (37a) with (14) and (25a) it is seen that (42) is equivalent to the particle equations (14) and (16a).

With γ_{ij} as the metric tensor and for the Newtonian coordinate system ordinary derivatives and covariant derivatives are equivalent. Hence, the covariant relation equivalent to (33) is

$$0=M_{i;k}{}^k-\tfrac{1}{2}g_{lm;i}h^{ln}M_n{}^m, \tag{38}$$

where the covariant derivatives are now defined with γ_{ij} as the metric tensor. Equation (38) constitutes 4 field equations for the 4 field quantities $\mathbf{v}$, ρ.

For a continuous medium the field equation, (30), for ϵ becomes for the Newtonian coordinate system

$$\nabla^2\epsilon-\epsilon^2\frac{\partial^2\epsilon}{\partial t^2}=\frac{1}{2}\left[\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)^2+\frac{1}{\epsilon}(\nabla\epsilon)^2\right]-\frac{k}{2}\frac{\rho(1+\epsilon^2v^2)}{(1-\epsilon^2v^2)}. \tag{39}$$

This field equation can be expressed in terms of the energy momentum tensor of the gravitational field (27a). Multiply (39) by $-\epsilon_{,i}/k\epsilon$ which form (27a) can then be written for the Newtonian coordinate system as

$$0=G_{i,k}{}^k-\frac{1}{2}\frac{\rho}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{1+\epsilon^2v^2}{(1-\epsilon^2v^2)^{\frac{1}{2}}}\frac{1}{\epsilon}\epsilon_{,i}. \tag{40}$$

From (34) this can be written in generalized coordinates as the tensor equation.

$$0=G_{i;k}{}^k+\tfrac{1}{2}g_{lm;i}h^{ln}M_n{}^m. \tag{41}$$

The covariant derivatives are of course based upon γ_{ij} as the metric tensor. The four field equations, (41), are equivalent as (40) divided by $\epsilon_{,i}$ is independent of the index i.

Conservation relations for the two fields in interaction are obtained by adding together (38) and (41) to obtain

$$0=(G_i{}^k+M_i{}^k)_{;k}. \tag{42}$$

For the Newtonian coordinate system this reduces to an ordinary divergence and defines conservation relations for momentum and energy in a manner similar to that described earlier for the gravitational field alone. For the Newtonian frame the integration of the ordinary divergence of the total energy momentum tensor over the volume bounded by two surfaces of constant time gives, after converting to a surface integral,

$$\int_{t_1}(G_i{}^4+M_i{}^4)d^3x=\int_{t_2}(G_i{}^4+M_i{}^4)d^3x. \tag{43}$$

Hence the total energy and momentum of the system is conserved.

Angular momentum and torque stress densities may be defined for the Newtonian coordinate system by defining the quantities (not tensor components)

$$L_{\alpha\beta}{}^j=x_\alpha(G_\beta{}^j+M_\beta{}^j)-x_\beta(G_\alpha{}^j+M_\alpha{}^j), \quad \alpha, \beta\neq 4. \tag{44}$$

Conservation of angular momentum takes the form of the divergence relation

$$0=L_{\alpha\beta,j}{}^j, \tag{45}$$

which in a manner similar to the derivation of (43) gives

$$\int_t L_{\alpha\beta}{}^4d^3x=\text{const}, \tag{46}$$

where $L_{\alpha\beta}{}^4$ are the three components of angular momentum density of the field.

CENTRAL FORCES

In order to evaluate the constant k, the field equation (39) will be solved for the case of static central symmetry. The resulting solution can be compared directly with the Schwarzschild solution of the corresponding problem of general relativity.

We consider a central spherically symmetrical source of gravitational field. For points outside the source the

scalar ϵ satisfies the equation,

$$\nabla^2\epsilon-\frac{1}{2\epsilon}(\nabla\epsilon)^2=0. \tag{47}$$

The solution to this equation for which $\epsilon\to 1$ asymptotically is

$$\epsilon=\left(1+\frac{a}{r}\right)^2, \tag{48}$$

where a is a positive constant which must be adjusted to join properly with the interior solution of (39).

The tensor g_{ij} is obtained by substituting (48) in (10). This may be compared directly with the Schwarzschild solution[5] in isotropic coordinates for which

$$g_{11}=g_{22}=g_{33}=-\left(1+\frac{G_0M_0}{2r}\right)^4, \quad g_{44}=\left(\frac{1-G_0M_0/2r}{1+G_0M_0/2r}\right)^2. \tag{49}$$

Neglecting $(G_0M_0/2r)^2$ compared with unity the two expressions are equal. At the surface of the sun $(G_0M_0/2r)^2$ has a value of 10^{-12}. It is doubtful that it will ever be possible to detect observationally the effect of a term this small. Thus, the two theories give essentially the same result for the planet and light trajectories in the solar system.

They also give essentially the same result for the gravitational red shift. An atom at the surface of the sun emits a photon of energy

$$\hbar\omega=\hbar\omega_0\epsilon^{-\frac{1}{2}}, \tag{50}$$

where $\hbar\omega_0$ is the energy of the photon emitted by a similar atom at great distance where $\epsilon=1$. The photon travels without shift in either energy or frequency (Newtonian units). Compared with the photon of a distant atom it appears to be shifted in frequency by an amount

$$\Delta\omega=\omega_0-\omega=\omega_0(1-\epsilon^{-\frac{1}{2}}), \tag{51}$$

$$\frac{\Delta\omega}{\omega_0}\simeq\frac{kM_0}{16\pi r}\simeq\frac{G_0M_0}{r}, \tag{52}$$

which is the usual approximate expression from general relativity for the red shift.

One of the fundamental assumptions of the theory introduced earlier concerned the ϵ dependence of the rest energy of a particle. The assumption $M=M_0\epsilon^{-\frac{1}{2}}$ can be justified on the grounds of the applicability of a limited principle of equivalence. If it be assumed that the gravitational red shift disappears for an experiment performed wholly within a freely falling elevator, it is found that the above relation for the rest energy of a particle is the only one compatible with this assumption.

[5] R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Oxford University Press, London, 1939), p. 205.

The gravitational red shift (52) is compensated by a Doppler effect resulting from the change in velocity of fall during the propagation time of the light from one atom to the other. It is also found that light travels in "straight" paths in the elevator, meter sticks being defined as "straight" but being for Newtonian units curved by just the right amount to make the light beam appear to be straight.

To evaluate the constant k, we consider first a general static solution to (39) in the weak field approximation. Matter is at rest, gravitational forces being balanced by other stresses not specified. These other stresses would also be sources of gravitational fields but may be neglected in the weak field limit. We assume that $\epsilon\to\epsilon_0$ asymptotically. For static solutions (39) becomes

$$\nabla^2\epsilon-\frac{1}{\epsilon}(\nabla\epsilon)^2=2\epsilon^{\frac{1}{2}}\nabla^2\epsilon^{\frac{1}{2}}=-\frac{k}{2}\rho. \tag{53}$$

This can be written as

$$\nabla^2\epsilon^{\frac{1}{2}}=-\frac{k}{4}\frac{\rho_0}{\epsilon}, \tag{54}$$

where ρ_0 is the matter energy density with $\epsilon=1$. In the weak field approximation ϵ can be replaced by the constant ϵ_0 on the right side of (54). Then the general solution with the proper asymptotic dependence is

$$\epsilon^{\frac{1}{2}}=\epsilon_0^{\frac{1}{2}}+\frac{k}{16\pi\epsilon_0}\int\frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3x'. \tag{55}$$

For a spherically symmetric source (55) becomes

$$\epsilon^{\frac{1}{2}}=\epsilon_0^{\frac{1}{2}}+\frac{kM_0}{16\pi\epsilon_0 r}, \tag{56}$$

where M_0 is the volume integral of ρ_0 over the region interior to the field point r. $M_0/\epsilon_0^{\frac{1}{2}}$ is the weak field approximation for the internal matter energy. If (56) is squared and compared with (48), (10), and (49) for $\epsilon_0=1$, k can be evaluated as

$$k=16\pi G_0. \tag{57}$$

One interesting question concerns the dependence upon ϵ_0 of the ratio of gravitational to electrical forces between matter in the weak field approximation. Consider two identical electrically charged static spherical mass distributions. The gravitational force acting on one of the masses due to the other is from (56) and (17)

$$F_g=\frac{M}{2\epsilon}\frac{\partial\epsilon}{\partial r}=-\frac{k\dot{M}_0^2}{16\pi\epsilon_0^2r^2}. \tag{58}$$

The electrostatic interaction is

$$F_e=-\frac{q^2}{\epsilon_0 r^2}. \tag{59}$$

The ratio of these two forces is

$$\frac{F_e}{F_g}=\frac{G_0\epsilon_0 M_0^2}{q^2}. \tag{60}$$

For fixed total matter M_0 and charge q, the ratio varies as ϵ_0. By measuring this ratio, one can in principle experimentally determine local $\epsilon=\epsilon_0$.

ELECTROMAGNETIC INTERACTIONS

To include the Maxwell field in the formalism in a proper way, the procedure outlined earlier is followed again. The Lagrangian density of the Lorentz invariant theory is written with c replaced by ϵ^{-1}. To this is added (29) the density of the gravitational field. The total is, for the Newtonian frame

$$L=-\sum_i M_i(1-\epsilon^2 v_i^2)^{\frac{1}{2}}\delta(\mathbf{r}-\mathbf{r}_i)+\sum_i e_i(\mathbf{A}\cdot\mathbf{v}_i-\varphi)\delta(\mathbf{r}-\mathbf{r}_i)$$
$$+\frac{1}{8\pi}\left(\epsilon E^2-\frac{1}{\epsilon}B^2\right)+\frac{1}{2k}\left[\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)^2-\frac{1}{\epsilon}(\nabla\epsilon)^2\right], \tag{61}$$

where

$$\mathbf{E}=-\frac{\partial\mathbf{A}}{\partial t}-\nabla\varphi,$$
$$\mathbf{B}=\nabla\times\mathbf{A}, \tag{62}$$

and the Lorentz condition

$$\nabla\cdot\mathbf{A}+\frac{\partial\varphi}{\partial t}=0, \tag{63}$$

is assumed to be satisfied.

The variation principle

$$\delta\int L d^4x=0, \tag{64}$$

gives as the equation of motion of the ith particle

$$\frac{d}{dt}\left[\frac{M_i\mathbf{v}_i\epsilon^2}{(1-\epsilon^2 v_i^2)^{\frac{1}{2}}}\right]=e_i(\mathbf{E}+\mathbf{v}_i\times\mathbf{B})$$
$$+\frac{1}{2}\frac{1+\epsilon^2 v_i^2}{(1-\epsilon^2 v_i^2)^{\frac{1}{2}}}M_i\frac{1}{\epsilon}\nabla\epsilon. \tag{65}$$

Here the first term on the right is the Lorentz force and the second term is the previously obtained gravitational force. The variation principle also gives the two Maxwell equations that are not identities.

$$\nabla\times\left(\frac{1}{\epsilon}\mathbf{B}\right)-\frac{\partial}{\partial t}(\epsilon\mathbf{E})=4\pi\sum_i e_i\mathbf{v}_i\delta(\mathbf{r}-\mathbf{r}_i),$$
$$\nabla\cdot(\epsilon\mathbf{E})=4\pi\sum_i e_i\delta(\mathbf{r}-\mathbf{r}_i). \tag{66}$$

It also gives the field equation for the gravitational field as

$$\nabla^2\epsilon-\epsilon^2\frac{\partial^2\epsilon}{\partial t^2}=-k\left\{\sum_i\frac{M_i(1+\epsilon^2 v_i^2)}{2(1-\epsilon^2 v_i^2)^{\frac{1}{2}}}\delta(\mathbf{r}-\mathbf{r}_i)\right.$$
$$\left.+\frac{1}{8\pi}\left(\epsilon E^2+\frac{1}{\epsilon}B^2\right)-\frac{1}{2k}\left[\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)^2+\frac{1}{\epsilon}(\nabla\epsilon)^2\right]\right\}. \tag{67}$$

In this coordinate system the electromagnetic energy density and the gravitational energy density serve as source terms for the generation of gravitational waves but they couple with opposite signs. Per unit energy, the electromagnetic field couples with twice the strength of a slowly moving particle. This is in agreement with the extra factor of two for the gravitational deflection of light.

On the right side of (67) the first term represents the total effect of the particle (at a distance) and the particle self-fields are to be omitted from the 2nd and 3rd terms. Presumably with a proper classical field theory of particles the self-fields of the particle (2nd and 3rd terms) would be included as sources of gravitational field. Because of the difference in sign of these two terms, the self-electric and gravitational energies could be each large compared with the total particle energy and still generate the proper gravitational field. There is, hence, a possibility within this framework of a charged particle of very small characteristic radius (gravitational radius) held together by gravitational forces. On the other hand, one has little reason to believe that a theory having its origin in macroscopic phenomena only should be valid at such small distances.

There is one piece of unfinished business regarding the form of (20). It was shown that without violating invariance arguments this term could be multiplied by an arbitrary function of ϵ. However, on physical grounds it would be expected that for the Newtonian coordinate system the third term on the right of 73 would, like the 2nd term, be an energy density. For this to be true the function of ϵ must be omitted.

The Lagrangian density (61) is easily written in generally covariant form by introducing the covariant four potential $A_i=(\mathbf{A},-\varphi)$ and the usual antisymmetric field tensor

$$F_{ik}=A_{i,k}-A_{k,i}. \tag{68}$$

In this notation, for the Newtonian frame

$$L=-\sum_i M_{i0}\frac{ds}{dt}\delta(\mathbf{r}-\mathbf{r}_i)+\sum_i e_i A_j\frac{dx^j}{dt}\delta(\mathbf{r}-\mathbf{r}_i)$$
$$-\frac{\epsilon}{16\pi}F_{ik}F_{rs}h^{ir}h^{ks}+\frac{1}{2k}h^{ij}\epsilon_{,i}\epsilon_{,j}. \tag{69}$$

As $(ds/dt)\delta(\mathbf{r}-\mathbf{r}_i)$ is assumed to transform like a scalar, this Lagrangian density is a scalar. The invariant form

of (64) is

$$\delta\int L\sqrt{-\gamma}d^4x=0. \qquad (70)$$

As previously, divergence difficulties with self-fields can be avoided by using a macroscopic description of matter. In the usual way an energy-momentum tensor for the electromagnetic field can be defined as

$$E_i{}^j=-\frac{\epsilon}{4\pi}F_{ik}F_{nm}h^{km}h^{nj}-\delta_i{}^j\frac{\epsilon}{16\pi}F_{mn}F_{rs}h^{mr}h^{ns}. \qquad (71)$$

In the Newtonian coordinate system the energy density is

$$E_4{}^4=\frac{1}{8\pi}\left(\epsilon E^2+\frac{1}{\epsilon}B^2\right). \qquad (72)$$

The momentum density is

$$-E_\alpha{}^4=\frac{\epsilon}{4\pi}(\mathbf{E}\times\mathbf{B})_\alpha,\quad \alpha=1,2,3. \qquad (73)$$

The Poynting vector is

$$E_4{}^\alpha=\frac{1}{4\pi\epsilon}(\mathbf{E}\times\mathbf{B})_\alpha=-\frac{1}{\epsilon^2}E_\alpha{}^4. \qquad (74)$$

The remainder of the terms constitute the Maxwell stress tensor.

For a continuous charge-current distribution the charge and current densities can be combined in a four vector which in the Newtonian coordinate system is

$$j^k=(\mathbf{j},\rho). \qquad (75)$$

Making use of (68), (62), and (75) Maxwell's equations (66) can be written for a general coordinate system as

$$(\epsilon F_{sr}h^{is}h^{rj})_{;j}=-4\pi j^i. \qquad (76)$$

For the Newtonian coordinate system the covariant divergence reduces to an ordinary divergence.

For the Newtonian coordinate system the ordinary divergence of the energy-momentum tensor of the electromagnetic field is from (71) and (76)

$$E_{i,j}{}^j=-F_{ik}j^k-E_4{}^4\frac{1}{\epsilon}\epsilon_{,i}, \qquad (77)$$

$$E_{\alpha,j}{}^j=\rho E_\alpha+(\mathbf{j}\times\mathbf{B})_\alpha-\frac{1}{8\pi}\left(\epsilon E^2+\frac{1}{\epsilon}B^2\right)\frac{1}{\epsilon}\epsilon_{,\alpha},$$

$$\alpha=1,2,3, \qquad (78)$$

$$E_{4,j}{}^j=-\mathbf{j}\cdot\mathbf{E}-\frac{1}{8\pi}\left(\epsilon E^2+\frac{1}{\epsilon}B^2\right)\frac{1}{\epsilon}\epsilon_{,4}. \qquad (79)$$

Equation (78) represents the force per unit volume exerted by the electromagnetic field. Equation (79) is the negative of the rate of energy transfer from electromagnetic to other forms of energy.

Equation (77) can be written in general covariant form by first rewriting the last term as a derivative of the Lagrangian density of the electromagnetic field. In generally covariant form (77) becomes

$$E_{i;j}{}^j=-F_{ik}j^k+\frac{\partial}{\partial\epsilon}\left(\frac{\epsilon}{16\pi}F_{ik}F_{rs}h^{ir}h^{ks}\right)\epsilon_{,i}. \qquad (80)$$

For a continuous matter distribution [see (39)], (67) becomes

$$\nabla^2\epsilon-\epsilon^2\frac{\partial^2\epsilon}{\partial t^2}=-k\left\{\frac{1}{2}\frac{\rho(1+\epsilon^2v^2)}{(1-\epsilon^2v^2)}+\frac{1}{8\pi}\left(\epsilon E^2+\frac{1}{\epsilon}B^2\right)\right.$$

$$\left.-\frac{1}{2k}\left[\epsilon\left(\frac{\partial\epsilon}{\partial t}\right)^2+\frac{1}{\epsilon}(\nabla\epsilon)^2\right]\right\}. \qquad (81)$$

If this equation is multiplied by $-\epsilon_{,i}/k\epsilon$ and compared with (27a) and (40) it can be written as

$$0=G_{i,k}{}^k-\tfrac{1}{2}\rho\frac{1+\epsilon^2v^2}{1-\epsilon^2v^2}\frac{1}{\epsilon}\epsilon_{,i}-\frac{1}{8\pi}\left(\epsilon E^2+\frac{1}{\epsilon}B^2\right)\frac{\epsilon_{,i}}{\epsilon}. \qquad (82)$$

Equation (82) can be written in generally covariant form by making use of (34)

$$0=G_{i;k}{}^k+\tfrac{1}{2}g_{lm;i}h^{lm}M_n{}^m+\frac{\partial}{\partial\epsilon}\left(\frac{\epsilon}{16\pi}F_{ik}F_{rs}h^{ir}h^{ks}\right)\epsilon_{,i}. \qquad (83)$$

For a continuous medium (65) can be written as

$$\frac{d}{dt}\left(\frac{\rho\epsilon^2\mathbf{v}}{1-\epsilon^2v^2}\right)=\rho_e\mathbf{E}+\mathbf{j}\times\mathbf{B}+\tfrac{1}{2}\rho\frac{1+\epsilon^2v^2}{1-\epsilon^2v^2}\frac{1}{\epsilon}\nabla\epsilon. \qquad (84)$$

Here ρ again represents the matter energy density in a co-moving coordinate system. Making use of (32), (33), (34), (37), (62), (68), and (75), (84) can be written

$$0=M_{i;k}{}^k-\tfrac{1}{2}g_{lm;i}h^{ln}M_n{}^m-F_{ik}j^k. \qquad (85)$$

This is obviously generally covariant.

Defining the total energy-momentum tensor as the sum of the tensors for the electromagnetic field, gravitational field and matter

$$T_i{}^k=E_i{}^k+G_i{}^k+M_i{}^k, \qquad (86)$$

the conservation of energy and momentum is expressed by

$$T_{i;k}{}^k=0. \qquad (87)$$

The vanishing of the covariant divergence of $T_i{}^k$ follows from (85), (80), and (83).

The fact that the covariant divergence reduces to an ordinary divergence for the Newtonian coordinate system suffices to define conservation laws. Also angular momentum and torque densities can be defined for the

Newtonian coordinate system as

$$L_{\alpha\beta}{}^{j}=x_{\alpha}T_{\beta}{}^{j}-x_{\beta}T_{\alpha}{}^{j},\quad \alpha,\beta=1,2,3. \tag{88}$$

The angular momentum of a closed system is conserved.

Although the physical picture of polarizable space was used as a guide in constructing the above theory, this interpretation is not fundamental to the formalism. The formalism has a structure of its own, that is independent of the philosophic interpretations. In this more general sense the theory is simply a scalar field theory of gravity based upon a flat space-time and local Lorentz invariance.

COSMOLOGY

The cosmological principle was taken to be a fundamental assumption of the theory. Namely, from any fixed position point of a Newtonian frame the universe is assumed to be on the average uniform. This implies that matter is on the average fixed in position relative to the Newtonian coordinate frame, for motion would introduce a lack of uniformity as seen by an observer located where the matter would be moving. In like manner the scalar field variable ϵ and matter density must be position independent.

If the lack of uniformity of matter and its random motion are ignored, the time dependence of ϵ can be computed from energy conservation. Equation (87) becomes for the Newtonian frame, remembering the position independence of the densities

$$0=T_{4,k}{}^{k}=T_{4,4}{}^{4}, \tag{89}$$

or

$$T_{4}{}^{4}=U\quad(\text{const}). \tag{90}$$

Assuming that electromagnetic energy can be ignored, from (86), (35), and (27), (90) can be written

$$U=\frac{1}{2k}\epsilon\left(\frac{d\epsilon}{dt}\right)^{2}+\frac{\rho_0}{\sqrt{\epsilon}}, \tag{91}$$

with ρ_0 constant.

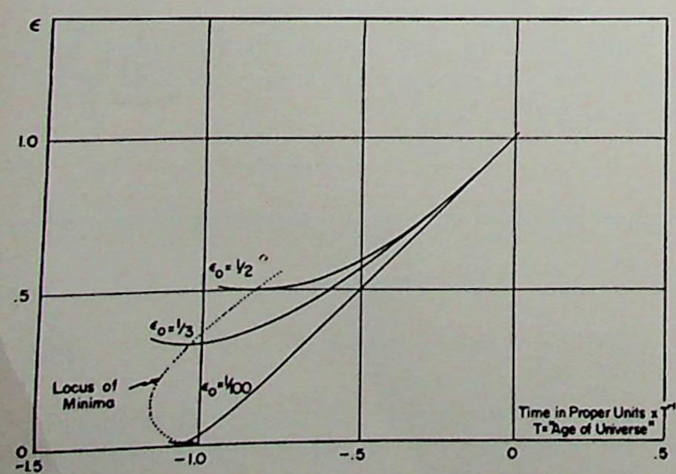


FIG. 1. Variation with time of vacuum polarizability for various energy densities.

Equation (91) can be integrated to give

$$\pm(2kU)^{\frac{1}{2}}t=2\epsilon^{\frac{1}{4}}(\epsilon^{\frac{1}{2}}-\epsilon_0{}^{\frac{1}{2}})^{\frac{1}{2}}\left[\tfrac{1}{3}\epsilon+(5/12)\epsilon_0{}^{\frac{1}{2}}\epsilon^{\frac{1}{2}}+\tfrac{5}{8}\epsilon_0\right]$$
$$+(5/4)\epsilon_0{}^{\frac{3}{2}}\log\left\{\left(\frac{\epsilon}{\epsilon_0}\right)^{\frac{1}{4}}+\left[\left(\frac{\epsilon}{\epsilon_0}\right)^{\frac{1}{2}}-1\right]^{\frac{1}{2}}\right\}, \tag{92}$$

with

$$\epsilon_0=\left(\frac{\rho_0}{U}\right)^{2}. \tag{93}$$

The constant of integration is chosen to cause the minimum of ϵ, (ϵ_0) to occur at $t=0$.

For some purposes it is desirable to obtain ϵ as a function of τ, the "proper" or atomic time. Multiplying (91) by

$$\left(\frac{dt}{d\tau}\right)^{2}=\epsilon, \tag{94}$$

and integrating gives

$$\pm(2kU)^{\frac{1}{2}}\tau=\epsilon^{\frac{1}{4}}(\epsilon^{\frac{1}{2}}-\epsilon_0{}^{\frac{1}{2}})^{\frac{1}{2}}(\epsilon^{\frac{1}{2}}+\tfrac{3}{2}\epsilon_0{}^{\frac{1}{2}})$$
$$+\tfrac{3}{2}\epsilon_0\log\left\{\left(\frac{\epsilon}{\epsilon_0}\right)^{\frac{1}{4}}+\left[\left(\frac{\epsilon}{\epsilon_0}\right)^{\frac{1}{2}}-1\right]^{\frac{1}{2}}\right\}. \tag{95}$$

Equation (95) with a shift in time zero is plotted in Fig. 1. U is chosen to give an asymptotic slope of unity and the time zero is adjusted to make $\epsilon=1$ at $t=0$.

Although all matter is at rest in this model there is a galactic red shift. With increasing ϵ, the photon emitted in the past has more energy than its present counterpart. This might be thought to cause a "blue shift." However, a photon loses energy with increasing ϵ at twice the rate of loss characteristic of an atom, hence there is a net shift toward the red. The energy lost by a photon becomes converted to gravitational field energy. Combining (16) and (16a), it is apparent that the energy of a photon varies as ϵ^{-1} whereas that of a stationary atom varies as $\epsilon^{-\frac{1}{2}}$. The red shift is consequently given by

$$\frac{\lambda_1}{\lambda_2}=\epsilon^{\frac{1}{2}}. \tag{96}$$

Here the ratio refers to the wavelength of light of distant origin λ_2 and the laboratory light λ_1. ϵ is the value of the dielectric constant at the time the light was emitted and ϵ is assumed to be unity now.

Although photon concepts were used to obtain the galactic red shift these particle ideas are not necessary. It is easily seen from Maxwell's equations that for time dependent but space independent ϵ, an electromagnetic wave propagates without a change in wavelength but with its frequency varying as ϵ^{-1}.

For small red shifts (96) can be written, assuming $\epsilon\gg\epsilon_0$

$$\frac{1}{\lambda}\frac{d\lambda}{d\tau}=\tfrac{1}{2}\epsilon^{-\frac{1}{2}}\frac{d\epsilon}{d\tau}=\tfrac{1}{2}(2kU)^{\frac{1}{2}}. \tag{97}$$

Using the present value of the Hubble constant of 6.5×10^9 years for the reciprocal of the left side of (97) gives an energy density of space

$$U=9\times10^{-29}\ \text{g/cm}^3. \tag{98}$$

Defining the "age of the universe" as the negative time at which ϵ extrapolates to zero (Fig. 1), gives an age of the universe of 3.25×10^9 years with this choice of U. Remembering the changes in the Hubble constant in the past years, another factor two change is perhaps not excluded. The evidence from isotope abundance and radioactive dating favors an age of the galaxy of at least 6.5×10^9 years.[6] If this is taken as the "age" of the universe, one obtains $U=2.25\times10^{-29}$ g/cm³. With a minimum matter density of the universe equal to that of galactic matter (10^{-30} g/cm³), there is a lower bound on ϵ_0 from (93) of

$$\epsilon_0>0.002. \tag{99}$$

It is interesting to consider the effect of a changing ϵ on the random motion in the universe. Consider a test particle moving through the idealized uniform space. Its momentum is a constant of the motion. Hence, from (15)

$$\epsilon^{\frac{1}{2}}\frac{\epsilon v}{(1-\epsilon^2v^2)^{\frac{1}{2}}}=\text{const.} \tag{100}$$

With increasing ϵ the ratio of particle to light velocity (ϵv) decreases. Thus, the random kinetic energy measured in atomic units falls with time and the kinetic temperature of galactic matter decreases with time. This energy is converted into gravitational energy [see (82)].

Much of the formalism developed in the past 40 years by the cosmologists is concerned with kinematics and is equally valid in the present theory. The tensor g_{ij} (10) is easily transformed into the standard form of Robertson and Lemaitre[7] through the substitution of proper time for coordinate time using (94). This gives as an expression for interval measured in proper units

$$ds^2=-\epsilon[(dx^1)^2+(dx^2)^2+(dx^3)^2]+d\tau^2, \tag{101}$$

with ϵ given by (104).

This is the type of universe characterized as open and flat. Since in the present theory the equations of motion of a particle are identical with those of general relativity, the kinematical description of the universe based upon (101) is identical with that of general relativity.

Equation (91) and its solution (92) are based upon the assumption that (37a) is valid or that the total matter is conserved. Only under this condition does the last term in (91) have this form. On the other hand, without violating conservation of energy, momentum, or angular momentum the condition (37a) could be relaxed. Namely, a mechanism for the production of particles by the conversion of gravitational energy into matter energy could be postulated. This would allow a variety of continuous creation universes.

A cataclysmic production of particles could also be obtained within the framework of the theory. The universe might be visualized as initially free of particles ($t<0$) and containing only gravitational energy. The time dependence of ϵ would be given by (92) with the negative sign and with $\epsilon_0=0$. ϵ would be decreasing with time, varying as $(-t)^{\frac{2}{3}}$. One might postulate in a completely *ad hoc* fashion the creation of heavy neutral bosons at a rate varying as $\epsilon^{-n}(d\epsilon/dt)^2$ with $n>\frac{1}{3}$. This would result in a cataclysmic production of bosons at $t=0$. The heavy bosons would then quickly decay into protons and electrons. The production of particles is accompanied by a rapid increase of ϵ from a low value to ϵ_0 after which production effectively ceases. From here on ϵ varies with time in accordance with (92).

THE PHYSICAL CONSTANTS OF NATURE

From (95) it is apparent that asymptotically ϵ is proportional to the age of the universe on an atomic time scale. From (53) and (95), the ratio of the gravitational to electrical interaction between two elementary particles varies asymptotically inversely as the age of the universe. This agrees with Dirac's hypothesis,[8] and suggests that the time dependence be examined of the remainder of the physical and astrophysical constants of Fig. 1 of the preceding paper.

The theory has been constructed to make the numbers in the first column constants, hence time independent. The present theory says nothing about the Fermi interactions except that their strength would suggest that they might be time dependent.

The only other number requiring discussion is in the last column, the number of particles in the universe out to the Hubble radius. Basing the discussion upon the asymptotic dependence expressed in atomic units, the volume inside the Hubble radius is proportional to $\tau^3\sim\epsilon^3$ and the average distance between particles varies as $\epsilon^{\frac{1}{2}}$. This gives a total number of particles which varies as $\tau^{\frac{3}{2}}$ instead of τ^2 as suggested by Dirac's considerations.

There is a simple explanation for this difference. The present theory contains a small dimensionless number ϵ/ϵ_0 which is time dependent. In the absence of a theory, any of the large dimensionless numbers could contain a factor in the form of a power of this number without changing its order of magnitude. Consequently, the time dependence of a number cannot be inferred from its magnitude alone.

From this point of view there is a single large dimensionless number which is statistical in origin. This is the number of particles in the universe. The age of the universe, "now," is not random but is conditioned by biological factors. The radiation rate of a star varies as

[6] F. Hoyle *et al.*, Science **124**, 611 (1956).

[7] See reference 5, p. 369 [Eq. (148.20)].

[8] P. A. M. Dirac, Proc. Roy. Soc. (London) **A165**, 199 (1938).

ϵ^{-7} [9] and for very much larger values of ϵ than the present value, all stars would be cold. This would preclude the existence of man to consider this problem. On the other hand, if ϵ/ϵ_0 were presently very much larger, the very rapid production of radiation at earlier times would have converted all hydrogen into heavier elements, again precluding the existence of man. This suggests that ϵ/ϵ_0 is presently a relatively small number, perhaps under ten. The universe can be characterized as young.

Some insight into the puzzle of the constants of Table I of the previous paper is gained by noting that for any solution to the field and orbit equations (39) and (14), a family of solutions can be obtained through the use of a simple scaling transformation.

Let n represent the density of particles of rest energy m. Then a macroscopic treatment of matter employs the matter density function ρ which is related to the particle density, according to (36a), as

$$nm=\frac{\rho}{(1-\epsilon^2v^2)^{\frac{1}{2}}}=\frac{\rho_0}{\sqrt{\epsilon}(1-\epsilon^2v^2)^{\frac{1}{2}}}. \qquad (102)$$

For any solution to (39), (14), and (102) another is obtained through the transformation

$$\begin{aligned}\epsilon(\mathbf{r},t)&\rightarrow\gamma^2\epsilon(\gamma^{-1}\mathbf{r},\gamma^{-3}t),\\ \mathbf{v}(\mathbf{r},t)&\rightarrow\gamma^{-2}v(\gamma^{-1}\mathbf{r},\gamma^{-3}t),\\ n(\mathbf{r},t)&\rightarrow\gamma n(\gamma^{-1}\mathbf{r},\gamma^{-3}t),\\ \rho_0(\mathbf{r},t)&\rightarrow\gamma\rho_0(\gamma^{-1}\mathbf{r},\gamma^{-3}t),\end{aligned} \qquad (103)$$

with γ a constant.

If this transformation is applied to the cosmological problem [and (92)] it is found that the number of particles in the universe scales as γ^4, the age and radius of the universe in atomic units scale as γ^2, and the ratio of electrical to gravitational forces scales as γ^2. The energy density U is independent of γ. Consequently this trans formation preserves the internal relations exhibited by Fig. 1 of the previous paper. Apparently, to make some sense out of the regularity of these numbers in the framework of the present theory, it is necessary to assume two things, that the universe is young (i.e., $\epsilon/\epsilon_0\sim1$) and that it has a characteristic energy density $\sim10^{-8}$ erg/cm^3 when expressed in Newtonian units.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the help of several of the author's colleagues, particularly Professor V. Bargmann who read the manuscript and made several helpful suggestions.

[9] E. Teller, Phys. Rev. 73, 801 (1948).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Dynamical Theory in Curved Spaces. I. A Review of the Classical and Quantum Action Principles*

BRYCE S. DEWITT

Department of Physics, University of North Carolina, Chapel Hill, North Carolina

INTRODUCTION

STUDY of the dynamical behavior of quantized physical systems, since the beginning of the quantum theory, has been almost exclusively restricted to systems satisfying either the requirements of nonrelativistic mechanics or, in the case of quantized fields, the requirements of special relativity. This paper is the first of a projected series, to appear in various journals, which will be devoted to extending the quantum theory to the basically broader context of general relativity.

The eventual program envisaged falls naturally into four parts: (1) Study and development of a number of usually neglected aspects of standard quantum mechanics which must be considered when attention is focused on the *nonlinear* features arising from the metric structure of the configuration space of a certain general class of systems. (2) Specialization to those members of this general class in which the specific covariant geometry of a four-dimensional continuum plays the leading role. (3) The carrying out of a reformulation of the "traditional" theory, in which specialized aspects of general relativity become prominently displayed. (4) A study of interacting systems, including those of the Fermi type which possess no classical analogs, in the context of general relativity.

In the following sections a start is made on the first of these four parts. The prototype for the general class of systems to be considered is the system consisting of a nonrelativistic particle moving in a curved space of n dimensions. The extension of standard quantum mechanics to this system is quite straightforward and involves nothing new. However, the notational developments required for the statement of well-known theorems in the context of curved spaces are generally unfamiliar. Therefore it was deemed worthwhile to begin at the beginning, and retrace familiar ground in considerable detail.

The motivation for the general program outlined is, of course, a desire ultimately to attack the problem of the role played by gravitation in the quantum domain. No apology will be made for this motivation, although needless to say, recent experiments have nothing to do with it![1] In the author's opinion it is sufficient that the problem is *there*, like the alpinist's mountain. Beyond that, however, the historical development of physics teaches a suggestive lesson in this connection, namely, that the existence of any fundamental theoretical structure which is far from having been pushed to its logical mathematical conclusions[2] is a situation which may have great potentialities.

The present work follows the traditional path in emphasizing classical-quantum analogies. At the same time an attempt is made to present the material from as modern a viewpoint as possible. In recent years a number of formal mathematical techniques have been developed which have a capacity for coming to grips with various physical problems in a more direct fashion than older methods, and which, owing to the additional insights which they therefore bestow, may eventually transform even the pedagogical approach to the quantum theory. Chief among these techniques are the quantization methods of Feynman[3] and Schwinger[4] and their various extensions, involving functional integration or differentiation and the analysis of the many characteristic Green's functions or "propagation functions" to which they give easy access. Since the problem of the quantum role of gravitation may justly be regarded as one of the most formidable in theoretical physics, it is well to have as many of these new methods available for use as possible. Accordingly, special attention is devoted in the following sections to the classical and quantum action principles, and a study of Feynman quantization in curved spaces is incorporated. The development presents no difficulty, but leads to a slight surprise, a rather curious ambiguity in the definition of the quantum Hamiltonian, involving the invariant curvature.

Extension of Schwinger's theory to curved spaces, on the other hand, is a different matter, and there is some doubt that it can be done at all, at least without extensive modification. Several independent serious efforts in this direction have thus far produced no success,[5]

* Work supported by the Institute of Field Physics.

[1] The total lack of pertinent experimental information is at once obvious if one recalls that the characteristic length for quantum-gravidynamical processes is $(\hbar G/c^3)^{\frac{1}{2}} \approx 10^{-33}$ cm (where G is the universal gravitation constant) corresponding to energies of the order of 10^{19} Bev.

[2] E.g., pre-Lorentzian electrodynamics, or pre-Lambian quantum electrodynamics.

[3] R. P. Feynman, Revs. Modern Phys. 20, 327 (1948).

[4] J. Schwinger, Phys. Rev. 82, 914 (1951). No attempt is made to compile a bibliography of the Feynman-Schwinger theory. Among the principal contributors mention may be made of Peierls, Edwards, Salam, Matthews, Källen, Lehmann, and Nambu.

[5] Private communication from P. G. Bergmann, and unpublished work of the author.

although the problem must presently be regarded as still an open one.

Subsequent papers will deal with the important problems posed by the existence of constraints, with systems obeying Fermi statistics in a nonlinear context, and finally with the actual quantization procedure for the gravitational field. An interesting fact will be noted in the course of this development, namely, that whenever a system possesses a classical analog, its quantum theory is in every case essentially completely determined by the corresponding classical theory. This may be regarded as the ultimate extension of Bohr's famous correspondence principle.

1. CLASSICAL TRANSFORMATION THEORY

A classical dynamical system is described by means of a set of $2n$ variables q^i, p_i, $i=1\cdots n$, which, if the system is subject to no constraints, may at a given initial instant be independently specified. The q^i are commonly known as *coordinates* and the p_i as *momenta*. The fact that n may be nondenumerably infinite for many systems of primary interest is of no importance in the present discussion. The simplicity of the formal mathematics for the finite case will be retained throughout, it being tacitly assumed (as is in fact the case) that the limiting procedures implied in the passage to continuum cases involve no difficulties of principle.

An experimental measurement performed on a physical system generally gives information about the value of some function(al) F of the q's and p's. F is known as a physical *observable*. In the classical theory any function(al) of the q's and p's is, in principle, an observable.

Of fundamental importance in the development of the dynamical theory is the *Poisson bracket* of two observables F and G, which is defined by

$$(F,G)\equiv\frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i}-\frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q^i}. \tag{1.1}$$

In this paper, a summation is to be understood over repeated indexes unless the contrary is explicitly stated. For many purposes it is convenient to introduce a slightly more compact notation through the replacements

$$x_i^1\equiv q^i,\quad x_i^2\equiv p_i. \tag{1.2}$$

Equation (1.1) may then be rewritten in the form

$$(F,G)\equiv\epsilon_{\alpha\beta}\frac{\partial F}{\partial x_i^\alpha}\frac{\partial G}{\partial x_i^\beta}, \tag{1.3}$$

where $\epsilon_{\alpha b}$ is the two-dimensional antisymmetric permutation symbol, the indexes α and β ranging over the values 1, 2.

The notation of (1.2) and (1.3) emphasizes the symmetry between coordinates and momenta which is characteristic of both classical and quantum transformation theory and which has its quantum expression in the wave-particle dualism. This symmetry does not, however, persist in detail in the dynamical theory of actual physical systems.[6] This fact has been anticipated here by the difference in the position of the index on the q's and p's.

From (1.1) or (1.3) the well-known corollaries follow immediately

$$(F_1,F_2)=-(F_2,F_1), \tag{1.4}$$

$$(F,f(F_1,F_2\cdots))=(F,F_1)\partial f/\partial F_1+(F,F_2)\partial f/\partial F_2+\cdots. \tag{1.5}$$

Also the Poisson-Jacobi identity

$$(F_1,(F_2,F_3))+(F_2,(F_3,F_1))+(F_3,(F_1,F_2))=0, \tag{1.6}$$

may be obtained with the aid of the three-dimensional permutation symbol ϵ_{abc} in terms of which the left side of (1.6) may be written

$$\tfrac{1}{2}\epsilon_{abc}(F_a,(F_b,F_c))=\epsilon_{abc}\epsilon_{\alpha\beta}\epsilon_{\gamma\delta}\frac{\partial F_a}{\partial x_i^\alpha}\frac{\partial F_b}{\partial x_j^\gamma}\frac{\partial^2 F_c}{\partial x_i^\beta\partial x_j^\delta}.$$

The quantity multiplying ϵ_{abc} in the latter expression is symmetric in a and b, and hence the whole expression vanishes.

Consider a transformation in the description of a dynamical system, from the variables q^i, p_i (or x_i^α) to a set of variables ξ_μ, $\mu=1\cdots 2n$. Consider the $2n$-dimensional antisymmetric matrix formed by taking the Poisson brackets of all possible pairs ξ_μ, ξ_ν. Using the notation of (1.3) it is not difficult to show that the determinant of this matrix is given by

$$|(\xi_\mu,\xi_\nu)|=[\partial(x)/\partial(\xi)]^2. \tag{1.7}$$

Evidently, if the Jacobian $\partial(x)/\partial(\xi)$ of the transformation $x_i^\alpha\rightarrow\xi_\mu$ is nonvanishing the matrix (ξ_μ,ξ_ν) has an inverse. This inverse defines the so-called *Lagrange bracket*[7] $[\![\xi_\mu,\xi_\nu]\!]$, satisfying

$$[\![\xi_\mu,\xi_\sigma]\!](\xi_\nu,\xi_\sigma)=\delta_{\mu\nu}. \tag{1.8}$$

Since

$$\epsilon_{\alpha\gamma}\epsilon_{\beta\gamma}=\delta_{\alpha\beta}, \tag{1.9}$$

the Lagrange bracket is given explicitly by

$$[\![\xi_\mu,\xi_\nu]\!]\equiv\epsilon_{\alpha\beta}\frac{\partial x_i^\alpha}{\partial\xi_\mu}\frac{\partial x_i^\beta}{\partial\xi_\nu}\equiv-[\![\xi_\nu,\xi_\mu]\!]. \tag{1.10}$$

Out of the set of all possible transformations q^i, $p_i\rightarrow\xi_\mu$ classical dynamics focuses attention on certain special ones known as *canonical transformations*. The set of all canonical transformations may be defined as

[6] Notwithstanding the fact that the symmetry idea has been sufficiently impressive to have been invoked on at least two occasions as a rigorous dynamical principle: M. Born, Proc. Roy. Soc. (London) **A165**, 291 (1938); H. Yukawa, Phys. Rev. **77**, 219 (1950).

[7] Whittaker [*Analytical Dynamics* (Cambridge University Press, New York, 1937), p. 298] uses the symbols ( ) and [ ] for the Poisson and Lagrange brackets, respectively. The symbol [ ] is, however, here reserved for the commutator bracket used in the quantum theory.

the most general set of transformations q^i, $p_i \to \bar{q}^i$, $\bar{p}_i$ (or $x_i^\alpha \to \bar{x}_i^\alpha$) such that

(i) there exists an essentially unique (up to trivial modifications) differential form of the type

$$A\bar{p}_i d\bar{q}^i + B\bar{q}^i d\bar{p}_i + C p_i dq^i + D q^i dp_i,$$

which is an exact differential for all members of the set.[8]

(ii) The members of the set form a group.

It must be specified that the coefficients A, B, C, D are independent of the p's and q's, and trivial cases may be excluded by imposing the conditions

$$A \neq B, \quad C \neq D. \tag{1.11}$$

Furthermore, the trivial exact differential

$$\tfrac{1}{2}(A+B)(\bar{p}_i d\bar{q}^i + \bar{q}^i d\bar{p}_i) + \tfrac{1}{2}(C+D)(p_i dq^i + q^i dp_i),$$

may be subtracted from the original form, and the result divided by $\frac{1}{2}(A-B)$ to obtain a canonical form

$$\bar{p}_i d\bar{q}^i - \bar{q}^i d\bar{p}_i + \zeta(p_i dq^i - q^i dp_i), \tag{1.12}$$

where $\zeta = (C-D)/(A-B)$. In order that a second transformation $\bar{q}^i$, $\bar{p}_i \to \tilde{q}^i$, $\tilde{p}_i$ be canonical the expression

$$\tilde{p}_i d\tilde{q}^i - \tilde{q}^i d\tilde{p}_i + \zeta(\bar{p}_i d\bar{q}^i - \bar{q}^i d\bar{p}_i), \tag{1.13}$$

must also be an exact differential. The group property of canonical transformations may then be applied to the result of subtracting ζ times (1.12) from (1.13), leading to the condition $\zeta^2 = -\zeta$, or

$$\zeta = -1. \tag{1.14}$$

Therefore, the set of all canonical transformations may be characterized by the statement

$$\bar{p}_i d\bar{q}^i - \bar{q}^i d\bar{p}_i - p_i dq^i + q^i dp_i = \epsilon_{\alpha\beta}(x_i^\alpha dx_i^\beta - \bar{x}_i^\alpha d\bar{x}_i^\beta) = dV, \tag{1.15}$$

for some V.

An explicit form for V may depend on any or all of the q^i, p_i, $\bar{q}^i$, $\bar{p}_i$. However, if V is regarded as a function of the $\bar{q}^i$, $\bar{p}_i$ (or $\bar{x}_i^\alpha$) alone then evidently

$$\frac{\partial V}{\partial \bar{x}_j^\beta} = \epsilon_{\gamma\delta} x_k^\gamma \frac{\partial x_k^\delta}{\partial \bar{x}_j^\beta} - \epsilon_{\alpha\beta}\bar{x}_j^\alpha, \tag{1.16}$$

$$\frac{\partial^2 V}{\partial \bar{x}_i^\alpha \partial \bar{x}_j^\beta} = [\![\bar{x}_i^\alpha, \bar{x}_j^\beta]\!] - \epsilon_{\alpha\beta}\delta_{ij} + \epsilon_{\gamma\delta} x_k^\gamma \frac{\partial^2 x_k^\delta}{\partial \bar{x}_i^\alpha \partial \bar{x}_j^\beta}. \tag{1.17}$$

The expression on the right of (1.17) can be symmetric under interchange of the indexes i, α with the indexes j, β if and only if

$$[\![\bar{x}_i^\alpha, \bar{x}_j^\beta]\!] = \epsilon_{\alpha\beta}\delta_{ij}, \tag{1.18}$$

and hence

$$(\bar{x}_i^\alpha, \bar{x}_j^\beta) = \epsilon_{\alpha\beta}\delta_{ij}. \tag{1.19}$$

Equation (1.18) or (1.19) may be used in place of (1.15) as an equally valid characterization of a canonical transformation.

The importance of canonical transformations lies in the existence of their well-known invariants.[9] For example, the Poisson bracket of two observables F and G remains unchanged if in (1.1) the q^i, p_i are replaced by $\bar{q}^i$, $\bar{p}_i$. This is readily shown with the aid of (1.19):

$$(F,G) \equiv \epsilon_{\alpha\beta} \frac{\partial F}{\partial x_i^\alpha}\frac{\partial G}{\partial x_i^\beta} = \epsilon_{\alpha\beta} \frac{\partial \bar{x}_j^\gamma}{\partial x_i^\alpha}\frac{\partial \bar{x}_k^\delta}{\partial x_i^\beta}\frac{\partial F}{\partial \bar{x}_j^\gamma}\frac{\partial G}{\partial \bar{x}_k^\delta}$$

$$= (\bar{x}_j^\gamma, \bar{x}_k^\delta)\frac{\partial F}{\partial \bar{x}_j^\gamma}\frac{\partial G}{\partial \bar{x}_k^\delta} = \overline{(F,G)}. \tag{1.20}$$

Lagrange brackets remain similarly invariant. For many purposes, therefore, the q^i, p_i may be replaced by the $\bar{q}^i$, $\bar{p}_i$ as equally valid *canonical* variables.

The explicit construction of canonical transformations is conveniently carried out with the aid of *generators*. Four standard types of generators are commonly considered, which may be denoted by $S_{\pm\pm}$, and which are related to the V of (1.15) by

$$S_{\pm\pm} = \tfrac{1}{2}(V \pm \bar{q}^i\bar{p}_i \pm q^i p_i). \tag{1.21}$$

Computation of $dS_{\pm\pm}$ with the aid of (1.15) shows that the solution of a set of simultaneous equations of any one of the following four types,

$$\bar{p}_i = \partial S_{++}/\partial\bar{q}^i, \qquad q^i = \partial S_{++}/\partial p_i, \tag{1.22a}$$

$$\bar{p}_i = \partial S_{+-}/\partial\bar{q}^i, \qquad p_i = -\partial S_{+-}/\partial q^i, \tag{1.22b}$$

$$\bar{q}^i = -\partial S_{-+}/\partial\bar{p}_i, \qquad q^i = \partial S_{-+}/\partial p_i, \tag{1.22c}$$

$$\bar{q}^i = -\partial S_{--}/\partial\bar{p}_i, \qquad p_i = -\partial S_{--}/\partial q^i, \tag{1.22d}$$

yields a canonical transformation.[10] Here S_{++} is an arbitrary function of the p_i and $\bar{q}^i$ alone, S_{+-} of the q^i and $\bar{q}^i$ alone, S_{-+} of the p_i and $\bar{p}_i$ alone, and S_{--} of the q^i and $\bar{p}_i$ alone.

Three elementary but important examples of canonical transformations may be obtained immediately with the aid of (1.22a):

(*i*) *The identity transformation:*

$$S_{++} = p_i\bar{q}^i, \quad \bar{q}^i = q^i, \quad \bar{p}_i = p_i. \tag{1.23}$$

(*ii*) *"Phase" transformations:*

$$S_{++} = p_i\bar{q}^i + \Phi(\bar{q}), \quad \bar{q}^i = q^i, \quad \bar{p}_i = p_i + \partial\Phi/\partial q^i. \tag{1.24}$$

(*iii*) *Point transformations:*

$$S_{++} = p_i q^i(\bar{q}), \quad \bar{q}^i = \bar{q}^i(q), \quad \bar{p}_i = p_j(\partial q^j/\partial\bar{q}^i). \tag{1.25}$$

[8] Here the $d\bar{q}^i$, $d\bar{p}_i$ are to be expressed in terms of the dq^i, dp_i or vice versa.

[9] Using (1.18) it is not hard to show that integrals of the form $\int dq^{i_1}dp_{i_1}\cdots dq^{i_m}dp_{i_m}$, $m=1\cdots n$, the summation being carried out in such a way that the indexes $i_1\cdots i_m$ are all different, are canonically invariant. The case $m=n$ yields the canonical invariance of the volume element in phase space, or the alternative fact that the Jacobian of a canonical transformation is equal to $+1$.

[10] The most general canonical transformation can be obtained through a combination of these four types.

Of special importance in the development of the dynamical theory to follow are the *infinitesimal* canonical transformations:

$$\bar{q}^i = q^i + \delta q^i, \quad \bar{p}_i = p_i + \delta p_i, \quad \text{or} \quad \bar{x}_i^\alpha = x_i^\alpha + \delta x_i^\alpha. \qquad (1.26)$$

The function V of (1.15) must in this case be an infinitesimal v satisfying

$$dv = -\epsilon_{\alpha\beta}(\delta x_i^\alpha dx_i^\beta + x_i^\alpha d\delta x_i^\beta). \qquad (1.27)$$

An *infinitesimal generator* for the transformation is given by

$$s = \tfrac{1}{2}(v + \epsilon_{\alpha\beta} x_i^\alpha \delta x_i^\beta), \qquad (1.28)$$

satisfying

$$ds = \epsilon_{\alpha\beta} dx_i^\alpha \delta x_i^\beta, \qquad (1.29)$$

so that

$$\partial s/\partial x_i^\alpha = \epsilon_{\alpha\beta}\delta x_i^\beta, \quad \text{or} \quad \delta x_i^\alpha = -\epsilon_{\alpha\beta}\partial s/\partial x_i^\beta, \qquad (1.30a)$$

s being expressed as a function of the x_i^α alone. In terms of q^i, p_i (1.30a) becomes

$$\delta q^i = -\partial s/\partial p_i, \quad \delta p_i = \partial s/\partial q^i, \qquad (1.30b)$$

$s=0$ gives the identity transformation; an infinitesimal phase transformation is given by $s=\varphi(q)$, where φ is an infinitesimal function of the coordinates; and an infinitesimal point transformation is given by

$$s = p_i\lambda^i(q), \quad \bar{q}^i = q^i - \lambda^i, \quad \bar{p}_i = p_i + p_j\partial\lambda^j/\partial q^i, \qquad (1.31)$$

where the λ^i are infinitesimal.

If F is a function(al) of the q^i, p_i with fixed functional form, then under the transformation (1.30) its value suffers a change of amount

$$\delta F = (\partial F/\partial x_i^\alpha)\delta x_i^\alpha = -(F,s). \qquad (1.32)$$

If, on the other hand, F is regarded as a fixed physical magnitude whose functional dependence on the canonical variables changes under (1.30), the amount of this change may be defined as

$$\bar{\delta}F(q,p) = \bar{F}(q,p) - F(q,p), \qquad (1.33)$$

where

$$\bar{F}(\bar{q},\bar{p}) = F(q,p). \qquad (1.34)$$

Evidently

$$\bar{\delta}F = -\delta\bar{F} = -\delta F = (F,s), \qquad (1.35)$$

correct to the first infinitesimal order.

This last result may be used to rederive the canonical invariance of Poisson brackets under infinitesimal transformations. One writes

$$\overline{(F,G)} = \overline{(\bar{F}(\bar{q},\bar{p}),\, \bar{G}(\bar{q},\bar{p}))}$$
$$= \overline{(F(\bar{q},\bar{p}) + \bar{\delta}F(\bar{q},\bar{p}),\, G(\bar{q},\bar{p}) + \bar{\delta}G(\bar{q},\bar{p}))}$$
$$= (F,G) + \delta(F,G) + (F,\bar{\delta}G) + (\bar{\delta}F,G)$$
$$= (F,G) + (s,(F,G)) + (F,(G,s)) + (G,(s,F)),$$

which leads to (1.20) in virtue of the Poisson-Jacobi identity (1.6).

Consider now a finite canonical transformation from a set of variables $x_i^\alpha(1)$ to a second set $x_i^\alpha(2)$. From (1.15) we have

$$\epsilon_{\alpha\beta}[x_i^\alpha(1)dx_i^\beta(1) - x_i^\alpha(2)dx_i^\beta(2)] = dV, \qquad (1.36)$$

where V is some function of the $x_i^\alpha(1)$ and $x_i^\alpha(2)$. Suppose the variables $x_i^\alpha(1)$ and $x_i^\alpha(2)$ are subjected to independent infinitesimal canonical transformations:

$$\bar{x}_i^\alpha(1) = x_i^\alpha(1) + \delta x_i^\alpha(1),$$
$$\delta x_i^\alpha(1) = -\epsilon_{\alpha\beta}\partial s(1)/\partial x_i^\beta(1), \qquad (1.37)$$

$$\bar{x}_i^\alpha(2) = x_i^\alpha(2) + \delta x_i^\alpha(2),$$
$$\delta x_i^\alpha(2) = -\epsilon_{\alpha\beta}\partial s(2)/\partial x_i^\beta(2). \qquad (1.38)$$

Then there must exist some new function $\bar{V}$ which differs infinitesimally from V, such that

$$\begin{aligned} d\bar{V} &= \epsilon_{\alpha\beta}[\bar{x}_i^\alpha(1)d\bar{x}_i^\beta(1) - \bar{x}_i^\alpha(2)d\bar{x}_i^\beta(2)] \\ &= dV + \epsilon_{\alpha\beta}[\delta x_i^\alpha(1)dx_i^\beta(1) + x_i^\alpha(1)d\delta x_i^\beta(1) \\ &\qquad - \delta x_i^\alpha(2)dx_i^\beta(2) - x_i^\alpha(2)d\delta x_i^\beta(2)] \\ &= dV + d\{2[s(2)-s(1)] \\ &\qquad + \epsilon_{\alpha\beta}[x_i^\alpha(1)\delta x_i^\beta(1) - x_i^\alpha(2)\delta x_i^\beta(2)]\}. \end{aligned} \qquad (1.39)$$

Evidently

$$\bar{V} = V + \delta V + \bar{\delta}V, \qquad (1.40)$$

where

$$\delta V = \epsilon_{\alpha\beta}[x_i^\alpha(1)\delta x_i^\beta(1) - x_i^\alpha(2)\delta x_i^\beta(2)], \qquad (1.41)$$

$$\bar{\delta}V = 2[s(2)-s(1)]. \qquad (1.42)$$

The infinitesimal $\bar{\delta}V$ represents a change in the functional *form* of V which is required to characterize the new canonical transformation $\bar{x}_i^\alpha(1)\to\bar{x}_i^\alpha(2)$, whereas δV simply represents the change in the value of the original function due to the changes $\delta x_i^\alpha(1)$, $\delta x_i^\alpha(2)$ in the values of the variables.

If the canonical transformation $x_i^\alpha(1)\to x_i^\alpha(2)$ is described by means of one of the generators $S_{\pm\pm}$, then it follows from (1.21) and (1.42) that

$$\bar{\delta}S_{\pm\pm} = s(2) - s(1), \qquad (1.43)$$

in which $s(1)$ and $s(2)$ are to be re-expressed in terms of the variables suitable to the generator in question. We therefore have

Theorem:

The variation in the functional form of the generator of a finite canonical transformation, due to independent infinitesimal canonical transformations of its arguments, is equal to the difference of the independent infinitesimal generators.

2. THE CLASSICAL ACTION PRINCIPLE

The "trajectory" of a physical system is described by a set of functions $q^i(t)$, $i=1\cdots n$, which determine the configuration of the system at any time t. Classical dynamical theory may be based on the following

Dynamical postulate:

The class of all trajectories $q(t)$ of a physical system is determined by the unfolding-in-time of a canonical transformation.

This implies the existence, in addition to the "coordinates" $q^i(t)$, of a set of momenta $p_i(t)$ which depend on the trajectory, or which, alternatively, may help to specify the trajectory.

The infinitesimal generator of the canonical transformation $q^i(t)$, $p_i(t) \to q^i(t+\delta t)$, $p_i(t+\delta t)$, where δt is an arbitrary infinitesimal "displacement" in time, will have the general form[11]

$$s = -H(q(t), p(t), t)\delta t, \tag{2.1}$$

where H is some explicit function of the coordinates and momenta, and possibly also of the time. H is known as the *Hamiltonian function(al)*, or, simply, the *Hamiltonian*, and its exact form depends on the system in question. Equations (1.30b) together with the relations $\delta q^i = \dot{q}^i \delta t$, $\delta p_i = \dot{p}_i \delta t$ lead immediately to the familiar Hamiltonian equations for the trajectory

$$\dot{q}^i = \partial H/\partial p_i, \quad \dot{p}_i = -\partial H/\partial q^i, \tag{2.2}$$

the dot denoting the total time derivative. From (2.2) it follows that the time rate of change of any physical observable F is given by

$$\dot{F} = (F, H) + \partial F/\partial t \tag{2.3}$$

[see (1.32)], where the partial derivative $\partial/\partial t$ is taken with respect to any explicit dependence on t which F may have. Since the Poisson bracket is canonically invariant it may be taken with respect to the canonical variables appropriate to the time in question as well as with respect to any other canonical variables. If F and G are any two physical observables the time rate of change of their Poisson bracket is given by

$$d(F,G)/dt = (\dot{F}, G) + (F, \dot{G}), \tag{2.4}$$

which follows from (2.3) together with the Poisson-Jacobi identity (1.6).

The finite canonical transformation which relates the coordinates and momenta at two different times, t' and t'', may generally be described by one or more of the four types of generators considered in (1.22). In the present context these equations become

$$p_i(t'') = \partial S_{++}/\partial q^i(t''), \quad q^i(t') = \partial S_{++}/\partial p_i(t'), \tag{2.5a}$$

$$p_i(t'') = \partial S_{+-}/\partial q^i(t''), \quad p_i(t') = -\partial S_{+-}/\partial q^i(t'), \tag{2.5b}$$

$$q^i(t'') = -\partial S_{-+}/\partial p_i(t''), \quad q^i(t') = \partial S_{-+}/\partial p_i(t'), \tag{2.5c}$$

$$q^i(t'') = -\partial S_{--}/\partial p_i(t''), \quad p_i(t') = -\partial S_{--}/\partial q^i(t'). \tag{2.5d}$$

Very often we wish to single out a particular value of one of the canonical variables. This is done by affixing one or more primes to the variable in question. The value thus indicated will be *independent* of the particular value of the time involved. Thus,

$$q'^i(t) = q'^i, \quad p'_i(t) = p'_i, \quad \text{independent of } t. \tag{2.6}$$

To avoid confusion in application of this convention, the following modification in notation will be introduced:

$$S_{++}(q''(t'')|p'(t')) = S(q'',t''|p',t'), \tag{2.7a}$$

$$S_{+-}(q''(t'')|q'(t')) = S(q'',t''|q',t'), \tag{2.7b}$$

$$S_{-+}(p''(t'')|p'(t')) = S(p'',t''|p',t'), \tag{2.7c}$$

$$S_{--}(p''(t'')|q'(t')) = S(p'',t''|q',t'). \tag{2.7d}$$

That is, the values of the canonical variables and the values of the time will be separately specified in the generating functions. On the other hand, the + and − signs will be dropped, as indicated, since the arguments themselves are sufficient to indicate which generator is meant.

The finite generators must be related in some way to the Hamiltonian which generates the infinitesimal displacements in time. This relation is readily determined with the aid of the theorem of the preceding section, (1.43). For definiteness, consider the function $S(q'',t''|p',t')$. Let $\delta q''^i$, $\delta t''$, $\delta p'_i$, $\delta t'$ be arbitrary infinitesimal variations in its arguments. The variations $\delta t''$, $\delta t'$ will give rise to a change in the *form* of this generator, considered as a function of the q''^i, p'_i. From (1.43) and (2.1), the amount of this change is $s''-s' = -H''\delta t'' + H'\delta t'$ where H' and H'' are the values of the Hamiltonian at the times t' and t'', respectively, for the particular trajectory specified by the arguments q'', t'', p', t'. The variations $\delta q''^i$, $\delta p'_i$, on the other hand, give rise to a change in *value*, which may be computed with the aid of (2.5a). The total variation in the generator is

$$\delta S(q'',t''|p',t') = p''_i \delta q''^i - H''\delta t'' + q'^i \delta p'_i + H'\delta t', \tag{2.8a}$$

where $p''_i = p_i(t'')$, $q'^i = q^i(t')$ for the trajectory in question. Similarly,

$$\delta S(q'',t''|q',t') = p''_i \delta q''^i - H''\delta t'' - p'_i \delta q'^i + H'\delta t', \tag{2.8b}$$

$$\delta S(p'',t''|p',t') = -q''^i \delta p''_i - H''\delta t'' + q'^i \delta p'_i + H'\delta t', \tag{2.8c}$$

$$\delta S(p'',t''|q',t') = -q''^i \delta p''_i - H''\delta t'' - p'_i \delta q'^i + H'\delta t'. \tag{2.8d}$$

By expressing all the quantities on the right of (2.8) in terms of the arguments appearing on the left, one is led to the *Hamilton-Jacobi equations* for the various generators. These are, respectively,

$$\left.\begin{aligned} \partial S/\partial t'' + H(q'', \partial S/\partial q'', t'') &= 0 \\ -\partial S/\partial t' + H(\partial S/\partial p', p', t') &= 0 \end{aligned}\right\} \tag{2.9a}$$

$$\left.\begin{aligned} \partial S/\partial t'' + H(q'', \partial S/\partial q'', t'') &= 0 \\ -\partial S/\partial t' + H(q', -\partial S/\partial q', t') &= 0 \end{aligned}\right\} \tag{2.9b}$$

$$\left.\begin{aligned} \partial S/\partial t'' + H(-\partial S/\partial p'', p'', t'') &= 0 \\ -\partial S/\partial t' + H(\partial S/\partial p', p', t') &= 0 \end{aligned}\right\} \tag{2.9c}$$

$$\left.\begin{aligned} \partial S/\partial t'' + H(-\partial S/\partial p'', p'', t'') &= 0 \\ -\partial S/\partial t' + H(q', -\partial S/\partial q', t') &= 0 \end{aligned}\right\}. \tag{2.9d}$$

[11] The minus sign is arbitrary and is chosen to correspond to convention.

In the integration of the Hamiltonian equations (2.2) a given trajectory is generally determined by specifying the q^i and p_i at a given "initial" time. Use of the generating functions, on the other hand, implies a different determination of the trajectory, namely, by specifying either the coordinates *or* the momenta (not both) at two different times. Actually, such a specification is not always possible. For example, if a representation is chosen (as is customary) in which the q^i are physical coordinates, the function $S(p'',t''|p',t')$ may not exist (e.g., in the case of the free particle). On the other hand, the function $S(q'',t''|q',t')$ may be multivalued (as in the case of the particle in a box) reflecting the fact that there may be more than one trajectory (the reflected trajectories) between the space-time points q', t' and q'', t''.

The function $S(q'',t''|q',t')$ is of special importance in the developments to follow, and is known as the *classical action*. We now study this function in greater detail, ignoring for the time being the possibility of its being multivalued. Let us denote by

$$q^i=(q'',t''|q^i(t)|q',t'), \tag{2.10}$$

the trajectory determined by the points q', t' and q'', t''. The explicit form of the right-hand side of (2.10) may be obtained by breaking into the trajectory at time t and solving the set of simultaneous equations

$$-\partial S(q'',t''|q,t)/\partial q^i=p_i(t)=\partial S(q,t|q',t')/\partial q^i. \tag{2.11}$$

The action S is so far uniquely determined only up to an arbitrary constant. In virtue of (2.11) and the Hamilton-Jacobi equations (2.9b), this constant may be so adjusted that for all q'', q, q', t'', t, t',

$$S(q'',t''|q',t')=-S(q',t'|q'',t''), \tag{2.12}$$

and

$$S(q'',t''|q',t')=[S(q'',t''|q,t)+S(q,t|q',t')]_{q=q_{\text{ext}}}, \tag{2.13}$$

the instruction "$q=q_{\text{ext}}$" indicating that the quantity in the brackets is to be made an extremal with respect to the q^i. The extremal values are given precisely by (2.10).

By breaking the trajectory into infinitely many pieces we see that the action may be regarded as formed in the following way:

$$S(q'',t''|q',t')=\{S[q]_{q',t'}{}^{q'',t''}\}_{q=(q'',t''|q(t)|q',t')}, \tag{2.14}$$

where $S[q]_{q',t'}{}^{q'',t''}$ is a functional of an arbitrary trajectory q between the points q', t' and q'', t'', satisfying the variational equation

$$\{\delta S[q]_{q',t'}{}^{q'',t''}/\delta q\}_{q=(q'',t''|q(t)|q',t')}=0, \tag{2.15}$$

and the combination laws

$$S[q]_{q',t'}{}^{q'',t''}=S[q]_{q(t),t}{}^{q'',t''}+S[q]_{q',t'}{}^{q(t),t}, \tag{2.16}$$

$$S[q]_{q',t'}{}^{q'',t''}=-S[q]_{q'',t''}{}^{q',t'}. \tag{2.17}$$

Equations (2.16) and (2.17) imply that $S[q]$ may be expressed in the form

$$S[q]_{q',t'}{}^{q'',t''}=\int_{q',t'}^{q'',t''}L(q,\dot{q},t)dt. \tag{2.18}$$

L is a function(al) of the q^i and their time derivatives (and also possibly of the time), known as the *Lagrangian* of the system in question.

When $S[q]$ is expressed in the form (2.18), the variational equation (2.15) is known as the *principle of stationary action*. Performing the indicated variation, with the restrictions $q^i(t')=q'^i$, $q^i(t'')=q''^i$, $\delta q^i(t')=0$, $\delta q^i(t'')=0$, and making an integration by parts, one obtains the familiar Lagrangian "equations of motion" for the trajectory

$$0=\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}. \tag{2.19}$$

Equations (2.19) are of the second order in contrast to the Hamiltonian equations (2.2) which are of the first order.

The relation between the Lagrangian and Hamiltonian functions may be determined with the aid of (2.8b), (2.14), and (2.19). Denote by $\delta q^i(t)$ the variation in a given trajectory produced by variations $\delta q'^i$, $\delta t'$, $\delta q''^i$, $\delta t''$ in its end points. Then, with an obvious notation,

$$\delta q^i(t')=\delta q'^i-\dot{q}'^i\delta t', \quad \delta q^i(t'')=\delta q''^i-\dot{q}''^i\delta t'', \tag{2.20}$$

and

$$\begin{aligned}
&\delta S(q'',t''|q',t')\\
&=\delta\left[\int_{t'}^{t''}Ldt\right]_{q=(q'',t''|q(t)|q',t')}\\
&=L''\delta t''-L'\delta t'\\
&\qquad+\left[\int_{t'}^{t''}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\delta q^i\right)dt\right]_{q=(q'',t''|q(t)|q',t')}\\
&=\left(\frac{\partial L}{\partial \dot{q}^i}\right)''\delta q''^i-\left[\left(\frac{\partial L}{\partial \dot{q}^i}\right)''\dot{q}''^i-L''\right]\delta t''\\
&\qquad-\left(\frac{\partial L}{\partial \dot{q}^i}\right)'\delta q'^i+\left[\left(\frac{\partial L}{\partial \dot{q}^i}\right)'\dot{q}'^i-L'\right]\delta t'.
\end{aligned} \tag{2.21}$$

Comparison of (2.8b) and (2.21) shows that

$$p_i=\partial L/\partial \dot{q}^i, \tag{2.22}$$

$$H=p_i\dot{q}^i-L. \tag{2.23}$$

A few tacit assumptions have been made. Firstly, it has been assumed that the Hamiltonian function, and hence the Lagrangian and action functions, are unique for a given system. Actually, a slight lack of uniqueness in these functions is allowed. The physics of a given

system is entirely contained in the Lagrange equations (2.19), and from their variational manner of derivation it is evident that they are unchanged if the Lagrangian is replaced by

$$\bar{L}=L+d\Phi/dt, \tag{2.24}$$

where Φ is an arbitrary function of the q^i and t. This replacement produces a change in the momenta, given by

$$\bar{p}_i=p_i+\partial\Phi/\partial q^i, \tag{2.25}$$

which is simply a phase transformation [(1.24)]. The corresponding transformation of the action is given by

$$\bar{S}[q]_{q',t'}{}^{q'',t''}=S[q]_{q',t'}{}^{q'',t''}+\Phi(q'',t'')-\Phi(q',t'), \tag{2.26}$$

$$\bar{S}(q'',t''|q',t')=S(q'',t''|q',t')+\Phi(q'',t'')-\Phi(q',t'), \tag{2.27}$$

which leaves the combination laws (2.16, 2.17) invariant. Phase transformations will be employed on several occasions in the future.

Secondly, the assumption that the generating function $S(q'',t''|q',t')$ exists at all (at least in some canonical representation) is equivalent to insisting that the first of the Hamiltonian equations (2.2) be solvable for the p_i in terms of the q^i, $\dot{q}^i$ (and possibly t). The Lagrangian equations (2.19) will then be obtained by substituting the resulting expressions for the p_i into the second of the Hamiltonian equations. If, however, the first of the Hamiltonian equations is not thus solvable the function $S(q'',t''|q',t')$ will not exist, and the Lagrangian function and the action principle will not exist in the usual sense. This case, the occurrence of which implies that the "velocities" $\dot{q}^i$ are not all independently specifiable, has been considered by Dirac[12] who has shown that a modified Lagrangian theory may still be constructed for it. It is of little practical importance, however, and will be ignored in this series of papers.

Much more important is the opposite case in which the momenta are not all independent, so that (2.22) cannot be solved to express the $\dot{q}^i$ in terms of the q^i, p_i (and t). This case has also been considered by Dirac[12] and leads to various possibilities, some of which will be carefully studied in subsequent papers of this series. One possible consequence, for example, is that the generating function $S(q'',t''|q',t')$, although it exists, may not define a unique trajectory [implying lack of unique solubility of (2.11)]. This would mean that many different extremal trajectories could be inserted into the right-hand side of (2.18), all leading to the same value of the action. In the present paper, however, such possibilities will be excluded.

3. CONSERVATION LAWS

Suppose we make a change in the representation of a physical system at any instant from a set of canonical variables $q^i(t)$, $p_i(t)$ to another set $\bar{q}^i(t)$, $\bar{p}_i(t)$. Consider the function V of (1.15) which characterizes this canonical transformation. If the form of V, considered as a function of q^i, p_i, $\bar{q}^i$, $\bar{p}_i$, remains constant in time then, by (1.42), the generators of infinitesimal displacements in time must be equal in the two representations. That is, $\bar{H}=H$ where H and $\bar{H}$ are the Hamiltonian functions for the two representations. (H and $\bar{H}$ may, of course, have different functional forms.) This result also follows from the canonical invariance of Poisson brackets, equations like (2.3) being independent of the choice of representation.

If, however, the form of V changes with time, so that V must be regarded as having an explicit dependence on t, H and $\bar{H}$ will no longer be equal but will be related by

$$-2(\bar{H}-H)\delta t=\bar{\delta}V. \tag{3.1}$$

If the change of representation is described by one of the generators $S_{\pm\pm}$ this equation may be replaced by

$$\bar{H}=H-\partial S_{\pm\pm}/\partial t, \tag{3.2}$$

which is a special case of the theorem at the end of Sec. 1.

It may happen that the generator is such that $\bar{H}=0$. From the right-hand side of (3.2) we see that this will be the case if the generator satisfies a Hamilton-Jacobi equation. The new variables $\bar{q}^i$, $\bar{p}_i$ are then constant in time and, when expressed as functions of the q^i, p_i, and t, are known as *constants of the motion*. It is clear that the functions $S(q'',t''|p',t')$, etc., considered in the previous section are special cases of such generators, which transform the canonical variables at an arbitrary time t'' back to their *constant* values at some fixed initial time t'.

A solution of the Hamilton-Jacobi equation gives a complete set of $2n$ independent constants of the motion. Except in very special cases, however, the Hamilton-Jacobi equation cannot be solved by finite methods, and is therefore unsuitable for the practical determination of constants of the motion. We must usually be content with a knowledge of only some of them, and for this purpose an investigation of infinitesimal changes of representation is more useful.

The change in the Hamiltonian under a canonical transformation generated by an infinitesimal s is given by

$$\bar{H}=H-\partial s/\partial t, \tag{3.3}$$

which follows from (1.28) and (3.1). Now suppose that $\bar{H}$, when expressed in terms of the $\bar{q}^i$, $\bar{p}_i$, has the same functional form as does H when expressed in terms of the q^i, p_i. Then, from (1.32),

$$H-(H,s)=H(\bar{q},\bar{p},t)=\bar{H}(\bar{q},\bar{p},t)=H-\partial s/\partial t,$$

or

$$\dot{s}=(s,H)+\partial s/\partial t=0. \tag{3.4}$$

Therefore, constants of the motion may be determined by discovering those infinitesimal canonical transforma-

[12] P. A. M. Dirac, Can. J. Math. 2, 129 (1950).

tions which leave the form of the Hamiltonian function invariant. A special case is provided by the Hamiltonian itself when it has no explicit dependence on the time. It then leaves itself invariant under its own infinitesimal transformations and is a constant of the motion.

Since the form of the Hamiltonian function remains invariant under an infinitesimal canonical transformation generated by a constant of the motion, the $\bar{q}^i$, $\bar{p}_i$ satisfy the same equations of motion as the q^i, p_i, and it is evident that a constant of the motion may also be regarded as the generator of an infinitesimal canonical transformation which transforms trajectories into trajectories.[13] From this point of view it is often easier to work directly from the Lagrangian than from the Hamiltonian. For example, suppose the Lagrangian function remains invariant under the transformation

$$\bar{q}^i = q^i - \delta q^i, \quad \dot{\bar{q}}^i = \dot{q}^i - \delta \dot{q}^i, \tag{3.5}$$

where the δq^i are certain infinitesimal functions of the q^i and t alone. Invariance of the Lagrangian function implies invariance of the Lagrangian equations and hence also of the Hamiltonian equations, so that $\bar{q}^i$ is a trajectory if q^i is. Moreover, invariances of this type are generally very easy to spot. Using the Lagrangian equations, we have

$$0 = \frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^i}\delta \dot{q}^i = \frac{d}{dt}(p_i \delta q^i), \tag{3.6}$$

so that the constant of the motion in this case is $p_i\delta q^i$. Evidently $p_i\delta q^i$ generates the transformation from the trajectory q^i to the trajectory $q^i - \delta q^i$.

If the Lagrangian function has no explicit dependence on the time then the classical action must remain invariant if the whole trajectory (together with its end points) is displaced in time by a constant δt as well as in coordinate space by the amounts δq^i above. Using previous results, we may express this by writing

$$0 = \delta S = p''_i \delta q''^i - p'_i \delta q'^i - (H'' - H')\delta t. \tag{3.7}$$

Since δt is arbitrary we have

$$H'' = H', \tag{3.8}$$

and

$$p''_i \delta q''^i = p'_i \delta q'^i. \tag{3.9}$$

Equations like (3.8) or (3.9) are known as *conservation laws.*

[13] This way of looking at things enables one to use a simple argument to prove that the Poisson bracket of two constants of the motion, F and G, is also a constant of the motion: If F and G are finite, render them infinitesimal through multiplication by a constant infinitesimal. Let the function G generate an infinitesimal canonical transformation which transforms trajectories into trajectories. The change in F under such a transformation from one trajectory to another is $\delta F = -(F,G)$. But since F, being a constant of the motion, is constant along both trajectories, its change δF is also a constant; q.e.d. This result can also be obtained formally from (2.4).

4. QUANTUM TRANSFORMATION THEORY

Historically the quantum theory has been built up by analogy from classical dynamics. Many routes may be followed in this development, and it is somewhat a matter of taste which is chosen. In order to parallel the development of Sec. 1 as closely as possible we begin with the construction of the quantum analog of the classical Poisson bracket.

Perhaps the most elegant formal argument leading to the quantum analog is that due to Dirac[14]: Let us assume, as given, the general mathematical framework of quantum mechanics, involving linear operators, state vectors, and their adjoints, and the probability interpretation of these quantities. Real physical observables are represented in the quantum theory by Hermitian operators F, G, etc. A quantum analog for the classical Poisson bracket is expected to satisfy identities paralleling as closely as possible (1.4, 1.5, 1.6). Equation (1.4) offers no particular difficulty. However, in generalizing (1.5) we must discover a rule for defining functions of several operators, and the derivatives of such functions. Formally we may expand the classical forms of these functions in series of products of their arguments. The classical arguments can then be replaced by the corresponding quantum operators, provided we can resolve the ambiguity which arises in the ordering of these operators due to the fact that they, unlike their classical counterparts, do not necessarily commute with one another.

In this series of papers we shall have to devote considerable attention to this *ordering problem.* However, let us set it aside for the present and suppose we have the quantum analog of some simple classical observable F expressed as a product of operators F_1, F_2, $\cdots$.

$$F = F_1 F_2 \cdots. \tag{4.1}$$

The differential law for this operator has the simple form

$$\delta F = [(\delta F_1)F_2\cdots] + [F_1(\delta F_2)\cdots] + \cdots. \tag{4.2}$$

Suppose now that the variation indicated is due to an infinitesimal canonical transformation, whatever that may mean in the present context. In the quantum theory as in the classical theory we try to express this variation in the form (1.32), the only difference being that s is now an infinitesimal operator. Thus

$$\delta F = -(F,s), \tag{4.3}$$

$$\delta F_i = -(F_i,s), \quad i = 1, 2 \cdots. \tag{4.4}$$

Equations (4.2, 4.3, 4.4) will be simultaneously satisfied for arbitrary s if and only if the following law is satisfied:

$$(F_1F_2, F_3) = (F_1,F_3)F_2 + F_1(F_2,F_3), \tag{4.5}$$

where F_1, F_2, F_3 are arbitrary operators, the order of

[14] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, New York, 1947), third edition, §21.

the factors being now important. Equation (4.5), combined with the identity

$$(F_1,F_2)=-(F_2,F_1), \tag{4.6}$$

implies also

$$(F_1,F_2F_3)=(F_1,F_2)F_3+F_2(F_1,F_3). \tag{4.7}$$

If one now considers a quantum Poisson bracket of the form (AB,CD) one finds that it may be re-expressed in two different ways with the aid of (4.5) and (4.7), depending on which is used first:

$$\begin{aligned}(AB,CD)&=[(A,C)D+C(A,D)]B\\&\quad+A[(B,C)D+C(B,D)] \quad (4.8a)\\&=[(A,C)B+A(B,C)]D\\&\quad+C[(A,D)B+A(B,D)]. \quad (4.8b)\end{aligned}$$

Subtracting one form from the other, one gets

$$0=(A,C)[B,D]-[A,C](B,D), \tag{4.9}$$

where [ ] denotes the commutator. Equation (4.9) can hold for arbitrary operators if and only if, for any pair A, B, (A,B) always stands in a constant numerical ratio to $[A,B]$. Since $[A,B]$ is anti-Hermitian if A and B are Hermitian, whereas (A,B) must be Hermitian, the numerical constant must be imaginary. Experimental physics tells us we must set

$$(A,B)=(i\hbar)^{-1}[A,B], \tag{4.10}$$

where $\hbar=1.054\times10^{-27}$ erg sec. From (4.10) it is but a step to show that quantum Poisson brackets also satisfy the Poisson-Jacobi identity:

$$\begin{aligned}&(F_1,(F_2,F_3))+(F_2,(F_3,F_1))+(F_3,(F_1,F_2))\\&=-\tfrac{1}{2}\hbar^{-2}\epsilon_{abc}(F_aF_bF_c+F_cF_bF_a\\&\qquad-F_aF_cF_b-F_bF_cF_a)=0. \quad (4.11)\end{aligned}$$

Combining (4.3) and (4.10) we have the familiar result that an infinitesimal canonical transformation is, in the quantum theory, simply an infinitesimal unitary transformation:

$$\bar{F}=F+\delta F=F+i\hbar^{-1}[F,s]=UFU^{-1}, \tag{4.12}$$

where

$$U=1-i\hbar^{-1}s, \tag{4.13}$$

$$U^{-1}=1+i\hbar^{-1}s. \tag{4.14}$$

More generally, a finite unitary transformation in the quantum theory will be expected to correspond in the classical theory to a finite canonical transformation. It should not be supposed, however, that the group of all unitary transformations in the quantum theory is isomorphic with the group of all canonical transformations in the classical theory. Such an isomorphism is prevented from existing by the ambiguity in the definition of the quantum analogs of general classical quantities, resulting from the noncommutativity of basic operators such as q^i, p_i. In general the most that can be established is a mapping from one group into the other, which becomes one-to-one only in the classical limit $\hbar\to0$. However, in the special cases to be considered in this series of papers, the ambiguity can be essentially removed from all physically significant quantum analogs, and a considerably stronger result can be established, namely, that there exists a *subgroup* of the full unitary group which *is* isomorphic to the group of all those canonical transformations which have any physical significance in the classical theory. This is an alternative statement of the fact that the classical theory in every case completely determines the corresponding quantum theory.

In the classical theory a canonical transformation can be regarded from two points of view, either as producing a change in the value of a physical observable F, as in (1.32), or as producing a change in the functional form of the observable, its value remaining fixed, as in (1.35). These alternatives have their analogs in the quantum theory. In the classical theory a change in the functional form is caused by a change of representation from one set of canonical variables to another. In the quantum theory the analogous change is in the *representative* $\langle\alpha''|F|\alpha'\rangle$ of F as the representation in terms of basic vectors $|\alpha'\rangle$ is changed. The $|\alpha'\rangle$ are orthogonal eigenvectors of a complete set of commuting Hermitian operators α_i. If the α_i are subjected to a unitary transformation

$$\bar{\alpha}_i=U\alpha_iU^{-1}, \tag{4.15}$$

the new operators $\bar{\alpha}_i$ will define an eigenbasis

$$|\bar{\alpha}'\rangle=U|\alpha'\rangle. \tag{4.16}$$

If U is given by (4.13) then the resulting change in the representative of F is

$$\begin{aligned}\bar{\delta}\langle\alpha''|F|\alpha'\rangle&=\langle\bar{\alpha}''|F|\bar{\alpha}'\rangle-\langle\alpha''|F|\alpha'\rangle\\&=\langle\alpha''|(U^{-1}FU-F)|\alpha'\rangle\\&=-i\hbar^{-1}\langle\alpha''|[F,s]|\alpha'\rangle\\&=-\langle\alpha''|\delta F|\alpha'\rangle.\end{aligned} \tag{4.17}$$

It is to be noted that the eigenvalues $\bar{\alpha}'$ and α' are identical. Therefore here, just as in our development of the classical theory, the use of primes will have the effect of singling out particular values of the variables, independently of the representation.

The generators of finite canonical transformations in the classical theory have also their quantum analogs, *viz.*, in the "transformation functions"

$$\begin{aligned}\langle\alpha''(2)|\alpha'(1)\rangle&=\langle\alpha''(1)|U^{-1}|\alpha'(1)\rangle \quad (4.18a)\\&=\langle\alpha''(2)|U^{-1}|\alpha'(2)\rangle \quad (4.18b)\end{aligned}$$

between two sets of commuting Hermitian operators, $\alpha_i(1)$ and $\alpha_i(2)$, which are related by a finite unitary

transformation

$$\alpha_i(2)=U\alpha_i(1)U^{-1}. \tag{4.19}$$

Suppose the operators $\alpha_i(1)$ and $\alpha_i(2)$ are subjected to independent infinitesimal canonical transformations:

$$\bar{\alpha}_i(1)=\alpha_i(1)+i\hbar^{-1}[\alpha_i(1),s(1)], \tag{4.20}$$

$$\bar{\alpha}_i(2)=\alpha_i(2)+i\hbar^{-1}[\alpha_i(2),s(2)]. \tag{4.21}$$

The corresponding changes in the eigenvectors $|\alpha'(1)\rangle$, $|\alpha'(2)\rangle$ will be

$$\bar{\delta}|\alpha'(1)\rangle=-i\hbar^{-1}s(1)|\alpha'(1)\rangle, \tag{4.22}$$

$$\bar{\delta}|\alpha'(2)\rangle=-i\hbar^{-1}s(2)|\alpha'(2)\rangle, \tag{4.23}$$

producing a change in the transformation function of amount

$$\bar{\delta}\langle\alpha''(2)|\alpha'(1)\rangle=i\hbar^{-1}\langle\alpha''(2)|[s(2)-s(1)]|\alpha'(1)\rangle. \tag{4.24}$$

Equation (4.24) is the quantum analog of the theorem at the end of Sec. 1.

In the problems usually considered (4.24) has an immediate application in the dynamical theory, just as (1.43) is applied in the classical theory to obtain the Hamilton-Jacobi equations. For the more general systems at which the present work is aimed, however, a slight modification of (4.24) will be necessary in order that the normalization of the vectors $|\alpha'(1)\rangle$ and $|\alpha'(2)\rangle$ be permitted to undergo desirable variations. We simply call attention to this fact here, postponing fuller discussion to the next section.

The question of normalization arises in the representation of an arbitrary state vector $|\psi\rangle$ as a superposition of eigenvectors $|\alpha'\rangle$, which is possible owing to the completeness condition on the operators α_i:

$$|\psi\rangle=\int|\alpha'\rangle d\alpha'\langle\alpha'|\psi\rangle, \tag{4.25}$$

where $\int d\alpha'$ indicates an integration and/or summation over all eigenvalues α'. The "volume element" $d\alpha'$ will generally contain a density function which serves to adjust the normalization. An explicit expression of the normalization is obtained by multiplying (4.25) on the left by $\langle\alpha''|$, which leads to

$$\langle\alpha''|\alpha'\rangle=\delta(\alpha'',\alpha'), \tag{4.26}$$

where

$$\int\delta(\alpha'',\alpha')f(\alpha')d\alpha'=f(\alpha''), \tag{4.27}$$

for arbitrary $f(\alpha')$.

5. QUANTUM DYNAMICS AND THE COORDINATE REPRESENTATION

Continuing the sequence of analogies to classical theory, one may base quantum dynamical theory on the

Dynamical postulate:

The temporal behavior of the operators representing the observables of a physical system is determined by the unfolding-in-time of a unitary transformation.

The infinitesimal generator of the unitary transformation $t\rightarrow t+\delta t$ will have the form $s=-H\delta t$ where H is the quantum analog of the classical Hamiltonian, assuming it can be unambiguously defined. The time rate of change of any physical observable will then be given by the quantum form of (2.3).

$$\dot{F}=(i\hbar)^{-1}[F,H]+\partial F/\partial t. \tag{5.1}$$

In order to discover the appropriate quantum form for a given Hamiltonian it will be necessary to study the basic operators out of which it is constructed, in particular the operators q^i, p_i corresponding to the classical coordinates and momenta. These operators must satisfy the commutation relations

$$[q^i,q^j]=0, \tag{5.2}$$

$$[q^i,p_j]=i\hbar\delta_j{}^i, \tag{5.3}$$

$$[p_i,p_j]=0, \tag{5.4}$$

corresponding to the analogous classical Poisson bracket relations. Equations (5.2) permit us to choose the q^i as a set of Hermitian operators defining a complete eigenbasis. A normal notation for expressing this would be

$$q^i(t)|q'(t)\rangle=q'^i(t)|q'(t)\rangle, \tag{5.5}$$

where the dependence of the q^i on the time is explicitly indicated. However, since the prime singles out a particular eigenvalue of q^i, independently of the time, it is more convenient to make the replacement

$$|q'(t)\rangle=|q',t\rangle, \tag{5.6}$$

and write

$$q^i|q',t\rangle=q'^i|q',t\rangle, \tag{5.7}$$

making the dependence of q^i on t explicit only when it is necessary to avoid confusion.

We now assume that a *natural metric* is defined in the space of the q^i, and choose the normalization of the $|q',t\rangle$ according to[15]

$$\langle q'',t|q',t\rangle=\delta(q'',q',t), \tag{5.8}$$

where

$$\int\delta(q'',q',t)f(q')d_tq'=f(q''), \tag{5.9}$$

for arbitrary $f(q')$, d_tq' being the invariant volume element. The notation reflects the fact that the metric, and hence the normalization, may change with time. The explicit form for d_tq' is

$$d_tq'=g^{\frac{1}{2}}(q',t)dq'^1\cdots dq'^n, \tag{5.10}$$

[15] The development here follows a previous paper: B. S. DeWitt, Phys. Rev. 85, 653 (1952).

where $g(q,t)$ is the determinant of the metric tensor. This implies

$$\delta(q'',q',t)=g^{-\frac{1}{2}}(q'',t)\delta(q''-q') \tag{5.11a}$$

$$=g^{-\frac{1}{2}}(q',t)\delta(q''-q'), \tag{5.11b}$$

$\delta(q''-q')$ being the ordinary n-dimensional delta function.

From (5.11) one may obtain the formal identities

$$(q''^i-q'^i)\delta(q'',q',t)=0, \tag{5.12}$$

$$\frac{\partial}{\partial q''^i}\delta(q'',q',t) =-\frac{\partial}{\partial q'^i}\delta(q'',q',t)-\frac{1}{2}\frac{\partial}{\partial q'^i}[\ln g(q',t)]\delta(q'',q',t), \tag{5.13}$$

$$(q''^i-q'^i)\frac{\partial}{\partial q''^j}\delta(q'',q',t)=-(q''^i-q'^i)\frac{\partial}{\partial q'^j}\delta(q'',q',t) =-\delta_j{}^i\delta(q'',q',t). \tag{5.14}$$

From these identities it is easy to obtain the representative of the operator p_i in the coordinate representation, the representative, or "matrix element," of q^i being

$$\langle q'',t|q^i|q',t\rangle=q'^i\delta(q'',q',t)=q''^i\delta(q'',q',t). \tag{5.15}$$

Taking the matrix element of (5.3), we get

$$i\hbar\delta_j{}^i\delta(q'',q',t)=\int(\langle q'',t|q^i|q''',t\rangle\langle q''',t|p_j|q',t\rangle -\langle q'',t|p_j|q''',t\rangle\langle q''',t|q^i|q',t\rangle)d_tq''' =(q''^i-q'^i)\langle q'',t|p_j|q',t\rangle, \tag{5.16}$$

which, in view of (5.12) and (5.14), implies

$$\langle q'',t|p_j|q',t\rangle =-i\hbar\frac{\partial}{\partial q''^j}\delta(q'',q',t)+F_j(q'',t)\delta(q'',q',t), \tag{5.17}$$

where the $F_j(q,t)$ are certain functions of the q^i and t. Alternative forms of (5.17) are

$$\langle q'',t|p_j|q',t\rangle =g^{-\frac{1}{2}}(q'',t)\left[-i\hbar\frac{\partial}{\partial q''^j}+G_j(q'',t)\right]\delta(q''-q') \tag{5.18a}$$

$$=g^{-\frac{1}{2}}(q'',t)\left[i\hbar\frac{\partial}{\partial q'^j}+G_j(q',t)\right]\delta(q''-q'), \tag{5.18b}$$

where

$$G_j=F_j+\tfrac{1}{2}i\hbar\partial(\ln g)/\partial q^j. \tag{5.19}$$

The commutation relations (5.4) will impose restrictions on the functions F_j. Taking the matrix element of (5.4) and using (5.18), one gets

$$0=\int(\langle q'',t|p_i|q''',t\rangle\langle q''',t|p_j|q',t\rangle -\langle q'',t|p_j|q''',t\rangle\langle q''',t|p_i|q',t\rangle)d_tq'''$$

$$=g^{-\frac{1}{2}}(q'',t)\Bigg[\hbar^2\frac{\partial^2}{\partial q''^i\partial q'^j}-i\hbar G_j(q',t)\frac{\partial}{\partial q''^i}+i\hbar G_i(q'',t)\frac{\partial}{\partial q'^j} +G_i(q'',t)G_j(q',t)-\hbar^2\frac{\partial^2}{\partial q''^j\partial q'^i}+i\hbar G_i(q',t)\frac{\partial}{\partial q''^j} -i\hbar G_j(q'',t)\frac{\partial}{\partial q'^i}-G_j(q'',t)G_i(q',t)\Bigg]\delta(q''-q')$$

$$=-i\hbar\left[\frac{\partial}{\partial q''^i}G_j(q'',t)-\frac{\partial}{\partial q''^j}G_i(q'',t)\right]\delta(q'',q',t)$$

$$=-i\hbar\left[\frac{\partial}{\partial q''^i}F_j(q'',t)-\frac{\partial}{\partial q''^j}F_i(q'',t)\right]\delta(q'',q',t) \tag{5.20}$$

which implies[16]

$$F_i=\partial F/\partial q^i \tag{5.21}$$

for some function F of the q^i and t.

A final restriction on F is imposed by the Hermitian condition on the p_i. We must have

$$\langle q'',t|p_i|q',t\rangle^*=\langle q',t|p_i|q'',t\rangle. \tag{5.22}$$

Insertion of (5.17) into (5.22) and use of the identity (5.13) yields the condition

$$F^\dagger=F+\tfrac{1}{2}i\hbar\ln g, \tag{5.23}$$

which implies that F has the form

$$F=-\Phi-\tfrac{1}{4}i\hbar\ln g, \tag{5.24}$$

where Φ is some real function of the q^i and t. The dependence of the representative of p_i on Φ may be removed by performing the unitary phase transformation

$$\bar{p}_i=e^{-(i/\hbar)\Phi}p_ie^{(i/\hbar)\Phi}=p_i+\partial\Phi/\partial q^i, \tag{5.25}$$

or alternatively by redefining the basis vectors according to

$$\overline{|q',t\rangle}=e^{(i/\hbar)\Phi}|q',t\rangle. \tag{5.26}$$

The transformation (5.26) simply changes the phase of the eigenvector $|q',t\rangle$ by the amount $\hbar^{-1}\Phi(q',t)$. Since this phase was arbitrary to begin with, we may assume that the adjustment (5.26) has already been made and write the representative of p_i in the standard form

$$\langle q'',t|p_i|q',t\rangle =-i\hbar\left\{\frac{\partial}{\partial q''^i}+\frac{1}{4}\left[\frac{\partial}{\partial q''^i}\ln g(q'',t)\right]\right\}\delta(q'',q',t). \tag{5.27}$$

[16] There is no difficulty in defining the derivative of an operator function of commuting variables.

A more conventional representation of the q^i, p_i is in terms of differential operators acting on the representative, or "wave function"

$$\psi(q',t)=\langle q',t|\psi\rangle \tag{5.28}$$

of an arbitrary state $|\psi\rangle$. The differential form $F_{q'}(t')$ of an operator F is defined by

$$F_{q'}(t')\psi(q',t')=\langle q',t'|F(t')|\psi\rangle, \tag{5.29}$$

and in particular,

$$q^i{}_{q'}=q'^i, \tag{5.30}$$

$$p_{iq'}=-i\hbar[\partial/\partial q'^i+\tfrac{1}{4}(\ln g')_{,i}]. \tag{5.31}$$

Here we use the abbreviations

$$f'\equiv f(q',t'), \tag{5.32}$$

$$f_{,i}\equiv \partial f/\partial q^i, \tag{5.33}$$

$$f'_{,i}\equiv \partial f(q',t')/\partial q'^i, \tag{5.34}$$

where f is an arbitrary function of the q^i and t. The prime is henceforth understood as occurring on the q^i and t together unless otherwise indicated. Repeated differentiation is denoted by additional indexes after the comma.

The assumption that coordinate space possesses a natural metric means that special importance is attached to point transformations in this space. We proceed to examine these transformations. Here we run into a slight notational problem. In the classical theory a point transformation is a canonical transformation, (1.25). In the quantum theory we may expect it to be describable as a unitary transformation:

$$\bar{q}^i=\bar{q}^i(q,t)=Uq^iU^{-1}. \tag{5.35}$$

If we now follow our usual convention, according to which a prime singles out a particular eigenvalue independently of the representation, we shall have $\bar{q}'^i=q'^i$ and $|\bar{q}',t'\rangle=U|q',t'\rangle$. But if we do this, a given eigenvalue will refer to different points of coordinate space, depending on the representation; $\bar{q}'$ will refer to that point whose coordinate values in the new representation are identical with the coordinate values of the original point in the original coordinate system. This raises difficulties in eigenvector normalization, since the normalization condition (5.8) and (5.9), depending on the invariant volume element, is designed to be independent of the coordinate system chosen. Therefore, it is more convenient to regard the prime as singling out a particular *point* of coordinate space[17] and write

$$\bar{q}'^i=\bar{q}^i(q',t'), \tag{5.36}$$

$$|\bar{q}',t'\rangle=|q',t'\rangle, \tag{5.37}$$

$$\bar{q}^i|q',t'\rangle=\bar{q}'^i|q',t'\rangle. \tag{5.38}$$

[17] This also avoids difficulties arising from coordinate transformations for which the spectra of q^i and $\bar{q}^i$ are not identical.

Under a point transformation the determinant of the metric tensor transforms according to

$$\bar{g}=|\partial q/\partial\bar{q}|^2 g, \tag{5.39}$$

where $|\partial q/\partial\bar{q}|$ denotes the coordinate Jacobian. Therefore

$$\ln\bar{g}=\ln g+2\ln|\partial q/\partial\bar{q}|, \tag{5.40}$$

and

$$\begin{aligned}\overline{(\ln g)}_{,i}&=\partial(\ln g)/\partial\bar{q}^i+2|\partial\bar{q}/\partial q|\partial|\partial q/\partial\bar{q}|/\partial\bar{q}^i\\ &=(\partial q^j/\partial\bar{q}^i)(\ln g)_{,j}+2(\partial q^j/\partial\bar{q}^i)_{,j}.\end{aligned} \tag{5.41}$$

Equation (5.41) may be used with (5.31) to obtain the point transformation law for the momenta:

$$\begin{aligned}\bar{p}_{iq'}&=-i\hbar[\partial/\partial\bar{q}'^i+\tfrac{1}{4}\overline{(\ln g')}_{,i}]\\ &=-i\hbar(\partial q'^j/\partial\bar{q}'^i)[\partial/\partial q'^j+\tfrac{1}{4}(\ln g')_{,j}]\\ &\qquad -\tfrac{1}{2}i\hbar(\partial q'^j/\partial\bar{q}'^i)_{,j}\\ &=\tfrac{1}{2}\{p_{jq'},\,\partial q'^j/\partial\bar{q}'^i\},\end{aligned} \tag{5.42}$$

or

$$\bar{p}_i=\tfrac{1}{2}\{p_j,\,\partial q^j/\partial\bar{q}^i\}, \tag{5.43}$$

where { } denotes the anticommutator. Equation (5.43) is the quantum generalization of the last of (1.25). The symmetrization effected by the anticommutator insures that the transformation leaves the Hermitian character of the p_i unchanged.[18]

An infinitesimal point transformation is given by

$$U=1-\tfrac{1}{2}i\hbar^{-1}\{p_i,\delta q^i\}, \tag{5.44}$$

$$\bar{q}^i=Uq^iU^{-1}=q^i-\delta q^i, \tag{5.45}$$

$$\bar{p}_i=Up_iU^{-1}=p_i+\tfrac{1}{2}\{p_j,\delta q^j{}_{,i}\}, \tag{5.46}$$

where the δq^i are arbitrary infinitesimal functions of the q's and t [see (1.31), (4.13)].[19] The vector $U(t')|q',t'\rangle$ is an eigenvector of $\bar{q}^i(t')$ corresponding to the eigenvalue q'^i. This eigenvalue refers, in the coordinate system $\bar{q}$, to a point which in the coordinate system q has the coordinate values $q'^i+\delta q'^i$. Therefore $U(t')|q',t'\rangle$ must be proportional to the eigenvector $|q'+\delta q',t'\rangle$:

$$U(t')|q',t'\rangle=A'^{\frac{1}{2}}|q'+\delta q',t'\rangle. \tag{5.47}$$

The proportionality constant $A'^{\frac{1}{2}}$ is not necessarily equal to unity since the operator U refers the normalization back to that obtaining at the point q', whereas the normalization required at the point $q'+\delta q'$ will generally be different. A' is readily evaluated by considering the

[18] Any other method of symmetrization would lead to the same result. For example, one might expand $\partial q^j/\partial\bar{q}^i$ in a power series in the q's. The operator p_j could then be inserted between the q's in any symmetrical fashion in each term of the series. The result of commuting p_j symmetrically to the left and to the right through the q's would be to produce two terms proportional to $\hbar$ which cancel each other, leaving (5.43).

[19] The unitary operator describing a finite point transformation has the general form $U=\exp[-\tfrac{1}{2}i\hbar^{-1}\{p_i,\Lambda^i\}]$ where the Λ^i are finite functions of the q's and t.

representative of an arbitrary state vector $|\psi\rangle$. We may write

$$
\begin{aligned}
A'^{\frac{1}{2}*}(\langle q',t'|\psi\rangle + \delta q'^i \partial\langle q',t'|\psi\rangle/\partial q'^i)
&= A'^{\frac{1}{2}*}\langle q'+\delta q',t'|\psi\rangle = \langle q',t'|U^{-1}(t')|\psi\rangle \\
&= \langle q',t'|\psi\rangle + \tfrac{1}{2}i\hbar^{-1}\langle q',t'|\{p_i,\delta q^i\}|\psi\rangle \\
&= \langle q',t'|\psi\rangle + \delta q'^i\partial\langle q',t'|\psi\rangle/\partial q'^i \\
&\quad + \tfrac{1}{2}[\delta q'^i{}_{,i} + \tfrac{1}{2}(\ln g')_{,i}\delta q'^i]\langle q',t'|\psi\rangle. \qquad (5.48)
\end{aligned}
$$

Comparison of the first and last lines shows that

$$A = 1 + \delta q^i{}_{.i}, \qquad (5.49)$$

where the dot followed by an index denotes the covariant derivative. The physical meaning of (5.49) is quite clear. If the points of any small region of coordinate space are subjected to displacements δq^i, the volume of the shifted region is increased by the factor A. Volume is conserved only if the displacements are divergenceless. This applies specifically to the displacement of the invariant volume element, which is effected by the operator U.

A similar situation exists in regard to displacements in time. The appropriate unitary operator in this case is

$$V = 1 + i\hbar^{-1}H\delta t, \qquad (5.50)$$

where δt is an arbitrary infinitesimal function of the time. The vector $V(t')|q',t'\rangle$ is an eigenvector of the operators $V(t')q^i(t')V^{-1}(t') = q^i(t'+\delta t')$ corresponding to the eigenvalues q'^i.[20] Therefore

$$V(t')|q',t'\rangle = B'^{\frac{1}{2}}|q',t'+\delta t'\rangle. \qquad (5.51)$$

The normalization constant is determined in this case by the representatives (5.15) and (5.27) of the q^i and p_i, respectively, and by the relation

$$
\begin{aligned}
\delta(q'',q',t+\delta t) &= \delta(q'',q',t) + \delta t\partial\delta(q'',q',t)/\partial t \\
&= [1 - \tfrac{1}{2}\delta t\partial\ln g(q'',t)/\partial t]\delta(q'',q',t) \qquad (5.52)
\end{aligned}
$$

which follows from (5.11). We have first of all

$$
\begin{aligned}
&\langle q'',t|q^i(t)|q',t\rangle \\
&= \langle q'',t|V^{-1}(t)V(t)q^i(t)V^{-i}(t)V(t)|q',t\rangle \\
&= B^{\frac{1}{2}*}(q'',t)B^{\frac{1}{2}}(q',t)\langle q'',t+\delta t|q^i(t+\delta t)|q',t+\delta t\rangle \\
&= B^{\frac{1}{2}*}(q'',t)B^{\frac{1}{2}}(q',t)q''^i\delta(q'',q',t+\delta t) \\
&= |B(q'',t)|[1 - \tfrac{1}{2}\delta t\partial\ln g(q'',t)/\partial t] \\
&\qquad \times\langle q'',t|q^i(t)|q',t\rangle, \qquad (5.53)
\end{aligned}
$$

which implies

$$B = 1 + [\tfrac{1}{2}\partial(\ln g)/\partial t + 2i\omega]\delta t, \qquad (5.54)$$

where ω is some real function of the q's and t. Secondly,

$$
\begin{aligned}
&\langle q'',t|p_i(t)|q',t\rangle \\
&= -i\hbar B^{\frac{1}{2}*}(q'',t)B^{\frac{1}{2}}(q',t) \\
&\qquad \times\left\{\frac{\partial}{\partial q''^i} + \frac{1}{4}\left[\frac{\partial}{\partial q''^i}\ln g(q'', t+\delta t)\right]\right\}\delta(q'', q', t+\delta t) \\
&= -i\hbar\left\{\frac{\partial}{\partial q''^i} + \frac{1}{4}\left[\frac{\partial}{\partial q''^i}\ln g(q'',t)\right]\right\} \\
&\quad \times|B(q',t)|\delta(q'', q', t+\delta t) \\
&\quad + i\hbar\left[\frac{\partial}{\partial q''^i}B^{\frac{1}{2}*}(q'',t) - \tfrac{1}{4}\delta t\frac{\partial^2\ln g(q'',t)}{\partial t\partial q''^i}\right]\delta(q'',q',t) \\
&= \langle q'',t|p_i(t)|q',t\rangle \\
&\qquad + \hbar\delta t[\partial\omega(q'',t)/\partial q''^i]\delta(q'',q',t), \qquad (5.55)
\end{aligned}
$$

which implies

$$\omega_{,i} = 0. \qquad (5.56)$$

We may set $\omega = 0$ by introducing a new representation

$$\overline{|q',t'\rangle} = \exp\left(i\int^{t'}\omega dt\right)|q',t'\rangle. \qquad (5.57)$$

Since ω is simply a numerical function of the time, this transformation leaves the representatives of the q^i, p_i unchanged, and we assume it already to have been carried out.

The change in the representative of an arbitrary state vector $|\psi\rangle$ under a displacement in time may then be expressed in the form

$$
\begin{aligned}
\langle q',t'+\delta t'|\psi\rangle &= B'^{-\frac{1}{2}}\langle q',t'|V^{-1}(t')|\psi\rangle \\
&= [1 - \tfrac{1}{4}(\partial\ln g'/\partial t')\delta t']\langle q',t'|\psi\rangle \\
&\quad + (i\hbar)^{-1}\langle q',t'|H(t')|\psi\rangle\delta t', \qquad (5.58)
\end{aligned}
$$

which leads to the Schrödinger equation for the wave function,

$$i\hbar\psi_{.t'}(q',t') = H_{q'}(t')\psi(q',t'), \qquad (5.59)$$

the dot followed by a t denoting *conservative* differentiation with respect to the time:

$$\psi_{.t} \equiv \partial\psi/\partial t + \tfrac{1}{4}(\partial\ln g/\partial t)\psi. \qquad (5.60)$$

When the metric is allowed to vary with time such a modified derivative is required in order to insure conservation of probability, as expressed by the time invariance of the integral

$$\int|\psi(q',t')|^2 d_{t'}q' = \langle\psi|\psi\rangle. \qquad (5.61)$$

We may finally write down the appropriate generalization of (4.24) for the case of the dynamical transformation function $\langle q'',t''|q',t'\rangle$. Using the symbol δ now

[20] These eigenvalues are the coordinate values of a point at the time $t'+\delta t'$. A point at the original time t' having the same coordinate values ought not, strictly speaking, to be called the *same* point, because (a) we have permitted the use of "moving" coordinate systems by allowing our point transformations to be time dependent, and (b) we have admitted the possibility that the metric structure of the coordinate space may vary with time. Under such circumstances it becomes somewhat meaningless to compare two points at different times.

to denote the operation

$$\delta\psi \equiv \psi_{.i}\delta q^i + \tfrac{1}{2}\psi\delta q^i{}_{.i} + \psi_{.t}\delta t, \tag{5.62}$$

we have

$$\delta\langle q'',t''|q',t'\rangle = i\hbar^{-1}\langle q'',t''|[\tfrac{1}{2}\{p_i(t''),\delta q^i(t'')\} - H(t'')\delta t'' - \tfrac{1}{2}\{p_i(t'),\delta q^i(t')\} + H(t')\delta t']|q',t'\rangle, \tag{5.63}$$

which has an obvious analogy with (2.8b).

It is possible to push the analogy even further by introducing a quantum analog for the classical action function. This is done with the aid of the concept of *well-ordered* operator functions.[21] Let $f(q'',t''|q',t')$ be an arbitrary function of the q'', t'', q', t'. The operator

$$f(q(t''),t''|q(t'),t') \equiv \int d_{t''}q'' \int d_{t'}q' |q'',t''\rangle f(q'',t''|q',t')\langle q',t'|, \tag{5.64}$$

is said to be the well-ordered operator form of $f(q'',t''|q',t')$. If $f(q'',t''|q',t')$ can be expanded as a power series in the q''^i, q'^i, its well-ordered operator form is obtained simply by replacing the q''^i, q'^i, respectively, by $q^i(t'')$, $q^i(t')$ in a *time-ordered* fashion. Any operator F may be expressed as a well-ordered operator $F(q(t''),t''|q(t'),t')$, the associated function being simply its matrix element:

$$F(q'',t''|q',t') = \langle q'',t''|F|q',t'\rangle. \tag{5.65}$$

In practice a given operator is generally expressed in a form which is not time ordered to begin with. To obtain it in its well-ordered form it is necessary to take into account the commutation effects of permuting its various components.

The quantum analog of the classical action is the well-ordered operator form of a function $\mathcal{S}(q'',t''|q',t')$ defined by

$$\langle q'',t''|q',t'\rangle = (g''g')^{-\frac{1}{4}} \exp[i\hbar^{-1}\mathcal{S}(q'',t''|q',t')]. \tag{5.66}$$

Using (5.62) one has

$$\begin{aligned}\delta\langle q'',t''|q',t'\rangle &= i\hbar^{-1}[(\partial\mathcal{S}/\partial q''^i)\delta q''^i - \tfrac{1}{2}i\hbar\delta q''^i{}_{,i} + (\partial\mathcal{S}/\partial t'')\delta t'' \\ &\quad + (\partial\mathcal{S}/\partial q'^i)\delta q'^i - \tfrac{1}{2}i\hbar\delta q'^i{}_{,i} + (\partial\mathcal{S}/\partial t')\delta t']\langle q'',t''|q',t'\rangle \\ &= i\partial^{-1}\langle q'',t''|[\delta\mathcal{S}(q(t''),t''|q(t'),t') \\ &\quad - \tfrac{1}{2}i\hbar\delta q^i{}_{,i}(t'') - \tfrac{1}{2}i\hbar\delta q^i{}_{,i}(t')]|q',t'\rangle.\end{aligned} \tag{5.67}$$

Comparison with (5.63) yields

$$\left.\begin{aligned} p_i(t'') &= \partial\mathcal{S}(q(t''),t''|q(t'),t')/\partial q^i(t''), \\ p_i(t') &= -\partial\mathcal{S}(q(t''),t''|q(t'),t')/\partial q^i(t'), \end{aligned}\right\} \tag{5.68}$$

$$\left.\begin{aligned} H(t'') &= -\partial\mathcal{S}(q(t''),t''|q(t'),t')/\partial t'', \\ H(t') &= \partial\mathcal{S}(q(t''),t''|q(t'),t')/\partial t', \end{aligned}\right\} \tag{5.69}$$

which are the analogs of the classical Hamilton-Jacobi equations. The derivatives on the right-hand sides must be taken with $\mathcal{S}$ in its well-ordered form.

The operator $\mathcal{S}$ is generally non-Hermitian, in spite of the fact that its variation $\delta\mathcal{S}$ *is* Hermitian when the displacements δq^i are chosen independent of the q's so as to commute with everything. In his development of the theory of quantized fields from a single dynamical principle Schwinger[4] assumes the existence of an Hermitian operator S which, when its components are ordered in some prescribed manner, possesses the same variation as $\mathcal{S}$ with q-independent δq^i. From the composition law

$$\begin{aligned}\delta\langle q'',t''|q',t'\rangle &= \int d_{t'''}q'''(\delta\langle q'',t''|q''',t'''\rangle\langle q''',t'''|q',t'\rangle \\ &\quad + \langle q'',t''|q''',t'''\rangle\delta\langle q''',t'''|q',t'\rangle) \\ &= i\hbar^{-1}\langle q'',t''|[\delta\mathcal{S}(q(t''),t''|q(t'''),t''') \\ &\quad + \delta\mathcal{S}(q(t'''),t'''|q(t'),t')]|q',t'\rangle.\end{aligned} \tag{5.70}$$

Schwinger infers that S may be chosen so as to satisfy[22]

$$S(t''|t') = S(t''|t''') + S(t'''|t'), \tag{5.71}$$

and hence may be expressed in the form

$$S(t''|t') = \int_{t'}^{t''} L dt, \tag{5.72}$$

where L is a quantum analog of the classical Lagrangian function. While such an operator certainly exists in the cases Schwinger considers, namely, interacting fields possessing a Lagrangian which never involves any one field more than quadratically, it is questionable, because of operator ordering complications, whether the same is true for the nonlinear systems of primary interest in the present investigation. The difficulties will be apparent after we have constructed the explicit Hamiltonian operator for a special case.

6. SPECIAL SYSTEM

As stated in the introduction, the prototype for the systems to be covered by the present analysis is the system consisting of a nonrelativistic particle moving in a curved space of n dimensions. The Lagrangian for this system has the general form

$$L \equiv \tfrac{1}{2}g_{ij}\dot{q}^i\dot{q}^j + a_i\dot{q}^i - v, \tag{6.1}$$

where the $g_{ij}(=g_{ji})$, a_i, v are functions of the q's and possibly also of the time t. In the case of an actual particle moving in three dimensions a_i and v may be the vector and scalar potentials, respectively, describing the effect of an impressed electromagnetic field. In general, however, a_i and v will not have such a specialized significance.

[21] P. A. M. Dirac, reference 14, §32. This famous section of Dirac's book has been phenomenally fruitful in its stimulation of the modern developments of the action principle.

[22] The operator $\mathcal{S}$ does *not* satisfy a relation of this type, although its variation does.

The form (6.1) embraces (at least approximately) all known physical systems which satisfy Bose statistics and possess a classical analog. It may be regarded as an invariant under point transformations,[23]

$$q^i = q^i(\bar{q}, t), \tag{6.2}$$

provided one imposes the transformation laws

$$\bar{g}_{ij} = \frac{\partial q^k}{\partial \bar{q}^i}\frac{\partial q^l}{\partial \bar{q}^j} g_{kl}, \tag{6.3}$$

$$\bar{a}_i = \frac{\partial q^j}{\partial \bar{q}^i}\left(a_j + g_{jk}\frac{\partial q^k}{\partial t}\right), \tag{6.4}$$

$$\bar{v} = v - a_i\frac{\partial q^i}{\partial t} - \tfrac{1}{2} g_{ij}\frac{\partial q^i}{\partial t}\frac{\partial q^j}{\partial t}. \tag{6.5}$$

The form is also invariant under phase transformations

$$\bar{L} = L + d\Phi/dt, \tag{6.6}$$

provided a_i and v are made to transform according to

$$\bar{a}_i = a_i + \Phi_{,i}, \tag{6.7}$$

$$\bar{v} = v - \partial\Phi/\partial t. \tag{6.8}$$

The equations of motion which follow from the Lagrangian function (6.1) are

$$q_{ij}\ddot{q}^j + [jk,i]\dot{q}^j\dot{q}^k + (\partial g_{ij}/\partial t - f_{ij})\dot{q}^j + \partial a_i/\partial t + v_{,i} = 0, \tag{6.9}$$

where

$$[jk,i] \equiv \tfrac{1}{2}(g_{ij,k} + g_{ik,j} - g_{jk,i}), \tag{6.10}$$

$$f_{ij} \equiv a_{j,i} - a_{i,j}. \tag{6.11}$$

These equations are evidently invariant in form under point and phase transformations. In the cases of greatest interest, which will eventually be studied in this series, there exist also other groups of transformations under which these equations remain invariant. When such other transformation groups exist the matrix (g_{ij}) is generally singular. We do not consider these cases in the present paper, but assume that (g_{ij}) is nonsingular and possesses an inverse (g^{ij}) satisfying

$$g^{ik}g_{kj} = \delta_j{}^i. \tag{6.12}$$

For a particle moving in a curved space, g_{ij} is the metric of the space (up to an arbitrary constant factor). We extend this interpretation and assume quite generally that g_{ij} is the natural metric for the space of the q^i, even when the q^i are field amplitudes, nondenumerably infinite in number. Thus we may therefore eventually be dealing with the Riemannian geometry of a space having a nondenumerable infinity of dimensions.

[23] If the point transformation is expressed in the form $\bar{q}^i = \bar{q}^i(q,t)$ then the derivatives $\partial q^i/\partial t$ may be replaced by $-(\partial q^i/\partial \bar{q}^j)\partial \bar{q}^j/\partial t$.

The momenta of the canonical formalism are given in the present case by

$$p_i = \partial L/\partial \dot{q}^i = g_{ij}\dot{q}^j + a_i. \tag{6.13}$$

Since (g_{ij}) is nonsingular these equations may be solved to express the "velocities" in terms of the momenta:

$$\dot{q}^i = g^{ij}(p_j - a_j). \tag{6.14}$$

The Hamiltonian then becomes

$$H = p_i\dot{q}^i - L = \tfrac{1}{2}g^{ij}(p_i - a_i)(p_j - a_j) + v. \tag{6.15}$$

The action function $S(q'', t'' | q', t')$ may be obtained either by solving the equations of motion (6.9) and substituting into the integral $\int L dt$, or by solving the Hamilton-Jacobi equations

$$\frac{\partial S}{\partial t''} + \tfrac{1}{2}g''^{ij}\left(\frac{\partial S}{\partial q''^i} - a''_i\right)\left(\frac{\partial S}{\partial q''^j} - a''_j\right) + v'' = 0, \tag{6.16}$$

$$-\frac{\partial S}{\partial t'} + \tfrac{1}{2}g'^{ij}\left(-\frac{\partial S}{\partial q'^i} - a'_i\right)\left(-\frac{\partial S}{\partial q'^j} - a'_j\right) + v' = 0. \tag{6.17}$$

We assume that the g_{ij}, a_i, v together with their first and second derivatives are continuous functions of the q^i and t. The quantities appearing in (6.16) may be expanded about the point q', t', giving

$$\begin{aligned} S(q'', t'' | q', t') &= (t''-t')^{-1}\{\tfrac{1}{2}g'_{ij}(q''^i - q'^i)(q''^j - q'^j) \\ &\quad + \tfrac{1}{12}(g'_{ij,k} + g'_{jk,i} + g'_{ki,j})(q''^i - q'^i)(q''^j - q'^j) \\ &\quad \times (q''^k - q'^k) + (1/72)[g'_{ij,kl} + g'_{ik,lj} + g'_{il,jk} \\ &\quad + g'_{kl,ij} + g'_{lj,ik} + g'_{jk,il} - g'^{mn}([ij,m]'[kl,n]' \\ &\quad + [ik,m]'[lj,n]' + [il,m]'[jk,n]')] \\ &\quad \times (q''^i - q'^i)(q''^j - q'^j)(q''^k - q'^k)(q''^l - q'^l) \\ &\quad + o(q''-q')^5\} + a'_i(q''^i - q'^i) \\ &\quad + \tfrac{1}{4}(a'_{i,j} + a'_{j,i} + \partial q'_{ij}/\partial t')(q''^i - q'^i)(q''^j - q'^j) \\ &\quad - (a'^i a'_i + v')(t''-t') + o(q''-q')^3 \\ &\quad + o[(t''-t')(q''-q')] + o(t''-t')^2, \end{aligned} \tag{6.18}$$

where $a^i = g^{ij}a_j$.

As a preliminary to the quantum-mechanical developments of the next section it is useful to examine the present system from a statistical viewpoint. Suppose we have an ensemble of a large number N of such systems, all identical, described by a distribution function $f(q,p,t)$ such that $Nf(q,p,t)dq^1\cdots dq^n dp_1\cdots dp_n$ represents the number of systems having coordinates in the range $dq^1\cdots dq^n$ at q and momenta in the range $dp_1\cdots dp_n$ at p, at the time t. Since $dq^1\cdots dq^n dp_1\cdots dp_n$ is a canonical invariant[9] and since the number of systems remains constant, $f(q,p,t)$ must have vanishing total time derivative. Written in the form

$$(f,H) + \partial f/\partial t = 0, \tag{6.19}$$

this statement becomes Liouville's conservation

theorem.[24] It may also be written in the integral form

$$\int f(q,p,t)dq^1\cdots dq^n dp_1\cdots dp_n=1. \quad (6.20)$$

Instead of asking for the number of systems in the range $dq^1\cdots dq^n dp_1\cdots dp_n$, and thereby specifying the trajectories of these systems by means of initial conditions q^i, p_i, one may also ask for the number of systems having trajectories specified by means of certain end-point conditions. Thus, one may work with a function $F(q'',t''|q',t')$ such that $NF(q'',t''|q',t')d_{t''}q''d_{t'}q'$ represents the number of systems having coordinates in the volume element $d_{t'}q'$ at q', t' and in the volume element $d_{t''}q''$ at q'', t''. From (2.8b) and (5.10),

$$F(q'',t''|q',t') = g''^{-\frac{1}{2}}D(q'',t''|q',t')g'^{-\frac{1}{2}}f(q',p'(q'',t''|q',t'),t'), \quad (6.21)$$

where

$$p'_i(q'',t''|q',t')=-\partial S(q'',t''|q',t')/\partial q'^i, \quad (6.22)$$

$$D=|D_{ji}|, \quad (6.23)$$

$$D_{ji}=\frac{\partial p'_i}{\partial q''^j}=-\frac{\partial^2 S}{\partial q''^j\partial q'^i}. \quad (6.24)$$

The determinant D, which is simply the Jacobian involved in changing the specification of the trajectory from the variables q', p' to the variables q', q'', was first introduced by Van Vleck.[25] It satisfies an important conservation law, namely,

$$\partial D/\partial t''+\partial(\dot{q}''^i D)/\partial q''^i=0, \quad (6.25)$$

$$\dot{q}''^i\equiv g''^{ij}(p''_j-a''_j)=g''^{ij}[(\partial S/\partial q''^j)-a''_j], \quad (6.26)$$

which follows from the insertion of (6.21) into the necessary conservation law

$$g''^{-\frac{1}{2}}\partial(g''^{\frac{1}{2}}F)/\partial t''+g''^{-\frac{1}{2}}\partial(g''^{\frac{1}{2}}\dot{q}''^i F)/\partial q''^i=0, \quad (6.27)$$

together with use of the identity

$$\partial p'_j/\partial t''+\dot{q}''^i\partial p'_j/\partial q''^i=0, \quad (6.28)$$

which can be obtained either by differentiating (6.16) with respect to q'^i or by observing that the p'_i may be regarded as constants of the motion, i.e., the *fixed* values of the momenta at an initial time t'. Equation (6.25) may also be derived directly, without appeal to the conservation law (6.27), by differentiating (6.28) with respect to q''^i, multiplying by D^{-1ji} where

$$D^{-1ik}D_{kj}=\delta_j{}^i, \quad (6.29)$$

and using the theorem

$$|A|^{-1}\delta|A|=\mathrm{tr}(A^{-1}\delta A), \quad (6.30)$$

which holds for an arbitrary nonsingular matrix A. In this way one may derive (6.27) instead of using it as a starting point.

If the Lagrangian (6.1) is regarded literally as describing a particle subject to forces derived from vector and scalar potentials in a curved n-dimensional space, then the distribution function $F(q'',t''|q',t')$ describes the motion of an ensemble of such particles. If $F(q'',t''|q',t')$ is nonvanishing only over a small range of the variables q'^i and q''^i, then the ensemble has the form of a cluster or "packet," the motion of which approximates that of a single particle. The density of particles in the packet is $N\rho''$ where

$$\rho''=\rho(q'',t'')=\int F(q'',t''|q',t')d_{t'}q', \quad (6.31)$$

satisfying the conservation laws

$$\int \rho'' d_{t''}q''=1, \quad (6.32)$$

$$g''^{-\frac{1}{2}}\partial(g''^{\frac{1}{2}}\rho'')/\partial t''+(\rho''\langle\dot{q}''^i\rangle_{\mathrm{Av}})_{.i}=0, \quad (6.33)$$

where the dot denotes covariant differentiation with respect to the q''^i, and where

$$\rho''\langle\dot{q}''^i\rangle_{\mathrm{Av}}=\int \dot{q}''^i F d_{t'}q'. \quad (6.34)$$

In the next section we shall need the first few terms in the explicit expansion of the Van Vleck determinant. Using the expansion (6.18) one finds

$$\begin{aligned}D_{ji}=(t''-t')^{-1}\{&g'_{ji}+[jk,i]'(q''^k-q'^k)\\&+\tfrac{1}{6}[[kl,i]'_{,j}+[lj,i]'_{,k}+[kj,i]'_{,l}\\&-g'^{mn}([ij,m]'[kl,n]'+[ik,m]'[lj,n]'\\&+[il,m]'[jk,n]')]\\&\times(q''^k-q'^k)(q''^l-q'^l)+o(q''-q')^3\}\\&+\tfrac{1}{2}(\partial g'_{ji}/\partial t'+f'_{ji})+o(q''-q')+o(t''-t'). \quad (6.35)\end{aligned}$$

Then, using the matrix theorem

$$\begin{aligned}|A+B|&=|A||1+A^{-1}B|\\&=|A|\{1+\mathrm{tr}(A^{-1}B)+\tfrac{1}{2}[\mathrm{tr}(A^{-1}B)]^2\\&\quad-\tfrac{1}{2}\,\mathrm{tr}(A^{-1}B)^2+\cdots\}, \quad (6.36)\end{aligned}$$

---

[24] Liouville, J. math. 3, 349 (1838). Any distribution function may evidently be used to generate an infinitesimal canonical transformation which changes trajectories into trajectories. The resulting change may be described as a "rotational" displacement of the phase-space trajectories in the regions where the distribution function is nonvanishing. Since an actual distribution function is never negative the "rotation" is always in the same sense.

[25] J. H. Van Vleck, Proc. Natl. Acad. Sci. 14, 178 (1928). This delightful paper extends the central idea of the WKB method and shows explicitly how quantum mechanics with its statistical interpretation passes asymptotically to the classical theory in the Correspondence Principle limit $\hbar\to 0$. Van Vleck considers the general case in which the variables q''^i, p''_i are referred back to an arbitrary canonical set α', β' and not merely to the q'^i, p'_i. If the α' are chosen as the action variables for a multiply-periodic conservative system, the asymptotic equivalence of the stationary state matrix elements of an arbitrary operator to the Fourier amplitudes of the corresponding classical quantity then follows immediately from his work.

which may be obtained by replacing A by $1+A^{-1}B$ in the integrated form of (6.30),

$$|A| = e^{\mathrm{tr}\ln A}, \tag{6.37}$$

and expanding both the logarithm and exponential, one finds

$$\begin{aligned} D(q'',t''|q',t') &= (t''-t')^{-n}g''^{\frac{1}{2}}g'^{\frac{1}{2}}\{1+\tfrac{1}{6}R'_{ij}(q''^i-q'^i)(q''^j-q'^j) \\ &\quad +o(q''-q')^3+o[(t''-t')(q''-q')] \\ &\quad +o(t''-t')^2\}, \end{aligned} \tag{6.38}$$

where

$$R_{ij} = -g^{kl}R_{ikjl}$$

$$\begin{aligned} R_{ikjl} &= \tfrac{1}{2}(g_{ij,kl}-g_{il,kj}-g_{kj,il}+g_{kl,ij}) \\ &\quad +g^{mn}([ij,m][kl,n]-[kj,m][il,n]). \end{aligned} \tag{6.39}$$

The R_{ikjl} are the components of the *Riemann tensor*. We have defined R_{ij} in such a way that the scalar

$$R = g^{ij}R_{ij} \tag{6.40}$$

is positive for a space of positive curvature.[26]

In (6.38) terms linear in $(q''^i-q'^i)$ and $(t''-t')$ inside the curly brackets have been removed by replacing the factor g', which would normally appear in front as a result of a straightforward application of (6.36) to (6.35), by the factor

$$\begin{aligned} g''^{\frac{1}{2}}g'^{\frac{1}{2}} &= g'[1+g'^{-\frac{1}{2}}(g'^{\frac{1}{2}})_{,i}(q''^i-q'^i) \\ &\quad +g'^{-\frac{1}{2}}(\partial g'^{\frac{1}{2}}/\partial t')(t''-t') \\ &\quad +\tfrac{1}{2}g'^{-\frac{1}{2}}(g'^{\frac{1}{2}})_{,ij}(q''^i-q'^i)(q''^j-q'^j)+\cdots], \end{aligned} \tag{6.41}$$

and noting that $g'^{-\frac{1}{2}}(g'^{\frac{1}{2}})_{,i} = \frac{1}{2}g'^{jk}g'_{jk,i}$, etc. The resulting more symmetric form for the Van Vleck determinant emphasizes the symmetry relation

$$D(q'',t''|q',t') = (-1)^nD(q',t'|q'',t''), \tag{6.42}$$

which is a necessary consequence of (2.12). A corollary of (6.25) and (6.42) is the conservation law for a packet moving backwards in time:

$$\partial D/\partial t' + \partial(\dot{q}'^iD)/\partial q'^i = 0. \tag{6.43}$$

7. FEYNMAN QUANTIZATION

In this section we adopt an approach to Feynman's theory due to Pauli,[27] which is based on Van Vleck's work.[25] We introduce the following structure,

$$\begin{aligned} \langle q'',t''|q',t'\rangle_c &= (2\pi i\hbar)^{-\frac{1}{2}n}g''^{-\frac{1}{4}}D^{\frac{1}{2}}(q'',t''|q',t')g'^{-\frac{1}{4}} \\ &\quad \times\exp[i\hbar^{-1}S(q'',t''|q',t')], \end{aligned} \tag{7.1}$$

and then show that it can be used to define the exact quantum transformation function in the limit $t''\to t'$. Expression (7.1) is an invariant under point transformations and, moreover, satisfies the relation

$$\langle q'',t''|q',t'\rangle_c^* = \langle q',t'|q'',t''\rangle_c \tag{7.2}$$

in virtue of (2.12) and (6.42). Next we show that

$$\lim_{t''\to t'}\langle q'',t''|q',t'\rangle_c = \delta(q'',q',t'). \tag{7.3}$$

To do this we first compare the action function $S(q'',t''|q',t')$ with the action function $S_0(q'',t''|q',t')$ for the same system with a_i and v set equal to zero and g_{ij} rendered time-independent. The expansion of $S_0(q'',t''|q',t')$ involves just the portion contained in the curly brackets on the right-hand side of (6.18). Therefore

$$\begin{aligned} S(q'',t''|q',t') &= S_0(q'',t''|q',t') \\ &\quad +o(q''-q')+o(t''-t'), \end{aligned} \tag{7.4}$$

$$D(q'',t''|q',t') = D_0(q'',t''|q',t')[1+o(t''-t')]. \tag{7.5}$$

Moreover, since the Hamiltonian function $H_0 = \frac{1}{2}g^{ij}p_{0i}p_{0j}$ is a constant of the motion for the modified system,

$$S_0(q'',t''|q',t') = \tfrac{1}{2}(t''-t')g'^{ij}p'_{0i}p'_{0j}. \tag{7.6}$$

Therefore, noting that as $t''\to t'$ the p'_{0i} become infinite except when $q''^i = q'^i$ (for all i), we may write, for an arbitrary function f,

$$\begin{aligned} &\lim_{t''\to t'}\int d_{t''}q''f(q'')\langle q'',t''|q',t'\rangle_c \\ &= \lim_{t''\to t'}(2\pi i\hbar)^{-\frac{1}{2}n}\int dp'_{01}\cdots dp'_{0n}f(q'')g''^{\frac{1}{4}} \\ &\qquad \times D_0^{-\frac{1}{2}}(q'',t''|q',t')g'^{-\frac{1}{4}}[1+o(q''-q')+o(t''-t')] \\ &\qquad \times\exp[i\hbar^{-1}S_0(q'',t''|q',t')] \\ &= [(t''-t')/2\pi i\hbar]^{\frac{1}{2}n}f(q')g'^{\frac{1}{2}} \\ &\qquad \times\int\exp[\tfrac{1}{2}i\hbar^{-1}(t''-t')g'^{ij}p'_{0i}p'_{0j}] \\ &\qquad \times dp'_{01}\cdots dp'_{0n} \\ &= f(q'), \end{aligned} \tag{7.7}$$

from which (7.3) follows.

The next step in the identification of $\langle q'',t''|q',t'\rangle_c$ is to show that it nearly satisfies a Schrödinger equation. Using (6.16), (6.25), and (6.38), one finds

$$\begin{aligned} &i\hbar\langle q'',t''|q',t'\rangle_{c,t''} - H_{q''}(t'')\langle q'',t''|q',t'\rangle_c \\ &= \tfrac{1}{2}\hbar^2g''^{-\frac{1}{4}}D^{-\frac{1}{2}}\partial[g''^{\frac{1}{2}}g''^{ij}\partial(g''^{-\frac{1}{4}}D)/\partial q''^j]/\partial q''^i \\ &\qquad \times\langle g'',t''|q',t'\rangle_c \\ &= [\tfrac{1}{12}\hbar^2R'+o(q''-q')+o(t''-t')]\langle q'',t''|q',t'\rangle_c, \end{aligned} \tag{7.8}$$

[26] This may be verified by direct computation for a spherical surface. In defining the Riemann tensor there are nearly as many conventions as authors. Expression (6.39) is, for example, the negative of the tensor defined by Bergmann [P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Inc., New York, 1946), Eq. (11.39)].

[27] W. Pauli, *Feldquantisierung*, lecture notes, Zürich (1950–1951), appendix.

where the operator $H_{q''}(t'')$ is defined by

$$H_{q''}(t'')\psi \equiv -\tfrac{1}{2}\hbar^2 g''^{-\frac{1}{2}}\frac{\partial}{\partial q''^i}\left(g''^{\frac{1}{2}}g''^{ij}\frac{\partial\psi}{\partial q''^j}\right)+i\hbar a''^i\frac{\partial\psi}{\partial q''^i}$$
$$+\left[\tfrac{1}{2}i\hbar g''^{-\frac{1}{2}}\frac{\partial}{\partial q''^i}(g''^{\frac{1}{2}}a''^i)+\tfrac{1}{2}a''^i a''_i+v''\right]\psi, \quad (7.9)$$

when acting on an arbitrary wave function $\psi(q'',t'')$. With the help of (5.31), which can be rewritten in the form

$$(g^{\frac{1}{4}}p_i g^{-\frac{1}{4}})_{q'} = -i\hbar\partial/\partial q'^i, \quad (7.10)$$

one sees that $H_{q''}(t'')$ is the differential form of the operator

$$H \equiv \tfrac{1}{2}g^{-\frac{1}{4}}(p_i - a_i)g^{\frac{1}{2}}g^{ij}(p_j-a_j)g^{-\frac{1}{4}}+v. \quad (7.11)$$

If it were not for the term on the right-hand side of (7.8) this operator could immediately be identified as the quantum analog of the Hamiltonian function (6.15) for the system (6.1). Because of this term, however, the quantum theory that one arrives at by applying the Correspondence Principle via (7.1) is determined *not* by the operator H but by the operator

$$H_+ \equiv H + \tfrac{1}{12}\hbar^2 R. \quad (7.12)$$

That is, in virtue of (7.3) and (7.8), one has

$$\lim_{t''\to t'}(t''-t')^{-1}[\langle q'',t''|q',t'\rangle_c - \langle q'',t''|q',t'\rangle_+]$$
$$=\lim_{t''\to t'}\frac{\partial}{\partial t''}[\langle q'',t''|q',t'\rangle_c - \langle q'',t''|q',t'\rangle_+]=0, \quad (7.13)$$

and hence

$$\langle q'',t''|q',t'\rangle_c = \langle q'',t''|q',t'\rangle_+ + o(t''-t')^2, \quad (7.14)$$

where $\langle q'',t''|q',t'\rangle_+$ is the transformation function generated by the Hamiltonian operator H_+. We shall discuss this curious result in further detail after developing the rest of the theory.

Let us first point out that the Schrödinger equation

$$i\hbar\psi_{,t'}(q',t') = H_{+q'}(t')\psi(q',t') \quad (7.15)$$

is invariant under point transformations. This is obvious if the point transformation is time independent, since the right-hand side of (7.9) involves only covariant derivatives and invariant combinations, and the extra term $\frac{1}{12}\hbar^2 R'$ in (7.15) is an invariant by itself. However, the conservative time derivative on the left of (7.15) is not invariant under time *dependent* point transformations $q^i = q^i(\bar{q},t)$. From (5.60) one has

$$\left.\begin{aligned}\psi_{.t}(q,t)&\equiv\left(\frac{\partial\psi}{\partial t}\right)_q+\frac{1}{4}\left(\frac{\partial \ln g}{\partial t}\right)_q\psi\\ \psi_{.t}(\bar{q},t)&\equiv\left(\frac{\partial\psi}{\partial t}\right)_{\bar{q}}+\frac{1}{4}\left(\frac{\partial \ln \bar{g}}{\partial t}\right)_{\bar{q}}\psi\end{aligned}\right\}. \quad (7.16)$$

Therefore, from the relation

$$\left(\frac{\partial\psi}{\partial t}\right)_{\bar{q}} = \left(\frac{\partial\psi}{\partial t}\right)_q + \frac{\partial\psi}{\partial q^i}\frac{\partial q^i}{\partial t}, \quad (7.17)$$

together with the transformation law (5.40) for $\ln g$, one may conclude that

$$\psi_{.t}(\bar{q},t) = \psi_{.t}(q,t)+\psi_{,i}\frac{\partial q^i}{\partial t}$$
$$+\frac{1}{4}\left[(\ln g)_{,i}\frac{\partial q^i}{\partial t}+2\frac{\partial\bar{q}^j}{\partial q^i}\frac{\partial^2 q^i}{\partial\bar{q}^j\partial t}\right]\psi. \quad (7.18)$$

The Schrödinger equation (7.15) nevertheless remains invariant, since, as one may readily verify, the transformation laws (6.4) and (6.5) for a_i and v add identical extra terms onto its right-hand side.

The Feynman formulation of the quantum theory follows from (7.14). If the transformation function is broken up into infinitely many pieces by means of the composition law

$$\langle q'',t''|q',t'\rangle_+$$
$$=\int\langle q'',t''|q''',t'''\rangle_+ d_{t'''}q'''\langle q''',t'''|q',t'\rangle_+, \quad (7.19)$$

one may replace $\langle\ \rangle_+$ by $\langle\ \rangle_c$ in each of the pieces and write

$$\langle q'',t''|q',t'\rangle_+ = \lim_{N\to\infty,\Delta t\to 0}\int d_{t^{(1)}}q^{(1)}\cdots$$
$$\times\int d_{t^{(N)}}q^{(N)}\langle q'',t''|q^{(N)},t^{(N)}\rangle_c\cdots$$
$$\times\langle q^{(2)},t^{(2)}|q^{(1)},t^{(1)}\rangle_c\langle q^{(1)},t^{(1)}|q',t'\rangle_c, \quad (7.20)$$

where $\Delta t = \max(t''-t^{(N)},\ \cdots t^{(2)}-t^{(1)},\ t^{(1)}-t')$, with $t''>t^{(N)}>\cdots>t^{(2)}>t^{(1)}>t'$. From the expression (7.1) for $\langle q'',t''|q',t'\rangle_c$ and the form of the expansion (6.18) for the action, the (in the limit) infinitely multiple integral receives significant contributions from the integrand only when the differences $q''^i - q^{(N)i}$, $\cdots q^{(2)i}-q^{(1)i}$, $q^{(1)i}-q'^i$ are of the order of $(\hbar\Delta t)^{\frac{1}{2}}$ or smaller. Therefore, introducing the symbol $\doteq$ to denote equivalence as far as use in the infinitely multiple integral is concerned, and using the easily verified relation

$$[2\pi i\hbar(t''-t')]^{-\frac{1}{2}n}\int d_{t''}q'' f(q'')(q''^i-q'^i)(q''^j-q'^j)$$
$$\times\exp[i\hbar^{-1}S(q'',t''|q',t')]$$
$$=i\hbar(t''-t')g'^{ij}f(q')+o(t''-t')^{\frac{3}{2}}, \quad (7.21)$$

one may write

$$\langle q'',t''|q',t'\rangle_c$$
$$\doteq[2\pi i\hbar(t''-t')]^{-\frac{1}{2}n}[1+\tfrac{1}{12}R'_{ij}(q''^i-q'^i)(q''^j-q'^j)]\times\exp[i\hbar^{-1}S(q'',t''|q',t')]$$
$$\doteq[2\pi i\hbar(t''-t')]^{-\frac{1}{2}n}[1+\tfrac{1}{12}i\hbar R'(t''-t')]\times\exp[i\hbar^{-1}S(q'',t''|q',t')]$$
$$\doteq[2\pi i\hbar(t''-t')]^{-\frac{1}{2}n}\exp[i\hbar^{-1}S_-(q'',t''|q',t')],\quad(7.22)$$

where S_- is the action function for a "classical" system with Hamiltonian function $H_-\equiv H-\frac{1}{12}\hbar^2R$ and Lagrangian function $L_-\equiv L+\frac{1}{12}\hbar^2R$:

$$S_-(q'',t''|q',t')=S(q'',t''|q',t')+\tfrac{1}{12}\hbar^2R'(t''-t')+o[(t''-t')(q''-q')].\quad(7.23)$$

The Feynman formulation now becomes

$$\langle q'',t''|q',t'\rangle_+=\int_{q',t'}^{q'',t''}\exp\left[i\hbar^{-1}\int_{t'}^{t''}L_-(q,\dot{q},t)dt\right]\delta[q],\quad(7.24)$$

where the symbols on the right-hand side are to be understood merely as formal abbreviations:

$$\int_{q',t'}^{q'',t''}\delta[q]\equiv\lim_{N\to\infty,\Delta t\to 0}\int\cdots\int[2\pi i\hbar(t''-t^{(N)})]^{-\frac{1}{2}n}d_{t^{(N)}}q^{(N)}\cdots\times[2\pi i\hbar(t^{(2)}-t^{(1)})]^{-\frac{1}{2}n}\times d_{t^{(1)}}q^{(1)}[2\pi i\hbar(t^{(1)}-t')]^{-\frac{1}{2}n},\quad(7.25)$$

$$\int_{t'}^{t''}L_-(q,\dot{q},t)dt\equiv\lim_{N\to\infty,\Delta t\to 0}[S_-(q'',t''|q^{(N)},t^{(N)})+\cdots+S_-(q^{(2)},t^{(2)}|q^{(1)},t^{(1)})+S_-(q^{(1)},t^{(1)}|q',t')].\quad(7.26)$$

The formal "functional integration" indicated in (7.24) has been described by Feynman[3] as a "path summation." According to his interpretation the transformation function $\langle q'',t''|q',t'\rangle_+$ is to be obtained by summing a complex amplitude over all possible paths in q, t space between the points q', t' and q'', t''. The phase of the amplitude associated with each path is obtained by integrating the classical Lagrangian along this path. In the case of systems for which the equations of motion are *linear* it is found that the specification of the path by means of intermediate points $q^{(1)}, t^{(1)}\cdots q^{(N)}, t^{(N)}$, together with the requirement that the path follow a true classical trajectory between these points [as implied by (7.26)] is actually unnecessary. The functional integral can be evaluated purely formally, without regard to its rigorous definition as an infinitely multiple integral. The result generally contains an unknown (infinite) normalization factor, which is properly accounted for when (7.25) is used but which always cancels out in the computation of any physically measurable quantity. The situation will certainly not be so simple in the general nonlinear cases involving curved spaces, because of the occurrence of a variable metric density in the volume elements $d_{t^{(1)}}q^{(1)}\cdots d_{t^{(N)}}q^{(N)}$.

The result expressed by (7.24) is even more curious than the previous result following from (7.8). In order to obtain the transformation function for a quantized system having Hamiltonian operator H_+ one must use, in the Feynman summation, the action corresponding to a classical system having Hamiltonian H_-. Whereas Pauli's formalism makes use of the Lagrangian L, Feynman must employ L_- to get the same result. When the author first discovered this phenomenon he thought that he had made an error in sign [e.g., in (7.21)] and that the occurrences of $\frac{1}{12}\hbar^2R$ actually cancel each other when one passes from Pauli's formalism on to that of Feynman. A straightforward computation patterned directly after Feynman's original paper,[3] however, demonstrates the reality of the phenomenon.[28]

Instead of stating the result in a symmetric manner one may also state it in the following forms: If the Lagrangian L is used in the Feynman summation then one obtains the transformation function $\langle q'',t''|q',t'\rangle_{++}$ generated by the Hamiltonian operator $H_{++}\equiv H+\frac{1}{6}\hbar^2R$. On the other hand, in order to obtain the transformation function $\langle q'',t''|q',t'\rangle$ generated by the operator H one must use the Lagrangian $L_{--}\equiv L+\frac{1}{6}\hbar^2R$ (corresponding to the Hamiltonian $H_{--}\equiv H-\frac{1}{6}\hbar^2R$) in the Feynman summation. If q space is flat then, of course, all these alternatives coalesce into a single theory. However, in the general case with nonvanishing R there is clearly an ambiguity in making the traditional passage from the classical theory to the quantum theory. The scalar curvature R is constructed out of quantities which already exist in the classical theory, and it has just the right dimensions, when multiplied by $\hbar^2$, to stand as a separate term in the Hamiltonian operator. The choice of a numerical factor to stand in front of $\hbar^2R$ is undetermined; all choices lead to the same classical theory in the limit $\hbar\to 0$. This, however, is the only ambiguity in the quantum Hamiltonian.

That the quantum theory goes over to the classical theory in the limit $\hbar\to 0$ is particularly transparent in the Feynman formalism. Thus, for example, in (7.24) when $\hbar$ is small the only paths which make in-phase

[28] J. L. Anderson (private communication).

contributions to the formal sum are those which cluster about that path between q', t' and q'', t'' which makes the action integral stationary, i.e., the classical trajectory. A wave packet $|\psi(q'',t'')|^2$ must move along such a trajectory, the particular trajectory taken depending on phase relationships in the initial wave function $\psi(q',t')$. The Pauli formalism also makes this obvious, for one may write $|\langle q'',t''|q',t'\rangle|^2 \approx |\langle q'',t''|q',t'\rangle_c|^2 = (2\pi\hbar)^{-n} g''^{-\frac{1}{2}} D(q'',t''|q',t') g'^{-\frac{1}{2}}$, and, in virtue of (6.25), this structure satisfies the classical packet conservation law (6.27), which in the quantum theory is simply the law of conservation of probability. This particular structure corresponds to an "exploding" packet, which initially has a delta-function form, with all values of momentum equally probable. However, as Van Vleck's work shows,[25] there is no need to make the initial specification in terms of coordinates q'^i; any complete set of commuting variables will do as well, and hence any type of packet may be specified.

The Feynman formalism provides a compact representation not only of the transformation function itself, but also of the matrix elements of arbitrary operators. For example, let f be an arbitrary function of the q's and t. Then, adopting the choice H_{++} for the Hamiltonian operator, which results from the use of the simple Lagrangian L in the path summation, we may write

$$\langle q'',t''|f(q(t'''),t''')|q',t'\rangle_{++}$$

$$= \int d_{t'''}q''' \langle q'',t''|q''',t'''\rangle_{++} f(q''',t''') \langle q''',t'''|q',t'\rangle_{++}$$

$$= \int_{q',t'}^{q'',t''} f(q(t'''),t''') \times \exp\left[i\hbar^{-1}\int_{t'}^{t''} L(q,\dot{q},t)dt\right]\delta[q]. \quad (7.27)$$

Here it is to be understood that q''', t''' is included among the variable points used to specify the path:

$$\int_{q',t'}^{q'',t''} \delta[q]$$

$$\equiv \lim_{M,N\to\infty,\Delta t\to 0} \int \cdots \int [2\pi i\hbar(t''-t^{(M+N)})]^{-\frac{1}{2}n}$$

$$\times d_{t^{(M+N)}}q^{(M+N)} \ldots [2\pi i\hbar(t^{(M+1)}-t''')]^{-\frac{1}{2}n}$$

$$\times d_{t'''}q'''[2\pi i\hbar(t'''-t^{(M)})]^{-\frac{1}{2}n}\ldots$$

$$\times d_{t^{(1)}}q^{(1)}[2\pi i\hbar(t^{(1)}-t')]^{-\frac{1}{2}n}. \quad (7.28)$$

Matrix elements involving the momentum may be obtained in a similar manner. For example,

$$\langle q'',t''|\tfrac{1}{2}\{f(q(t'''),t'''),p_i(t''')\}|q',t'\rangle_{++}$$

$$= \frac{1}{2}\int d_{t'''}q'''[\langle q'',t''|p_i(t''')|q''',t'''\rangle_{++}$$

$$\times f(q''',t''')\langle q''',t'''|q',t'\rangle_{++} + \langle q'',t''|q''',t'''\rangle_{++}$$

$$\times f(q''',t''')\langle q''',t'''|p_i(t''')|q',t'\rangle_{++}]$$

$$= \frac{1}{2}\int_{q',t'}^{q'',t''} f(q(t'''),t''')[\overleftarrow{p}_i(t''') + \overrightarrow{p}_i(t''')]$$

$$\times \exp\left[i\hbar^{-1}\int_{t'}^{t''} L(q,\dot{q},t)dt\right]\delta[q]. \quad (7.29)$$

Here one makes use of the coordinate representation (5.27) of the momentum operator, and the identity (5.13). The significance of the arrows in the final expression is obvious. The momentum at the point q''', t''' may be defined either with respect to the trajectory coming from the past or with respect to the trajectory going into the future:

$$\overleftarrow{p}'''_i = -\partial S(q^{(M+1)},t^{(M+1)}|q''',t''')/\partial q'''^i, \quad (7.30)$$

$$\overrightarrow{p}'''_i = \partial S(q''',t'''|q^{(M)},t^{(M)})/\partial q'''^i. \quad (7.31)$$

To obtain the symmetrized form involving the anticommutator, one must average the two.

A final example, in which the momenta appear quadratically, is given by

$$\langle q'',t''|g^{-\frac{1}{4}}(t''')p_i(t''')g^{\frac{1}{2}}(t''')f^{ij}(t''')p_j(t''')g^{-\frac{1}{4}}(t''')|q',t'\rangle_{++}$$

$$= -\hbar^2\int d_{t'''}q'''\langle q'',t''|q''',t'''\rangle_{++}$$

$$\times (f'''^{ij}\langle q''',t'''|q',t'\rangle_{++.j})_{.i}$$

$$= \hbar^2\int d_{t'''}q''' f'''^{ij}\langle q'',t''|q''',t'''\rangle_{++.i}\langle q''',t'''|q',t'\rangle_{++.j}$$

$$= \int_{q',t'}^{q'',t''} \overleftarrow{p}_i(t''') f^{ij}(q(t'''),t''')\overrightarrow{p}_j(t''')$$

$$\times \exp\left[i\hbar^{-1}\int_{t'}^{t''} L(q,\dot{q},t)dt\right]\delta[q]. \quad (7.32)$$

Here a covariant integration by parts has been performed,[29] the dot followed by one or more indexes denoting covariant differentiation with respect to the q'''^i.

It is now possible to derive Schwinger's dynamical principle directly from Feynman's formulation. Remembering that Schwinger's theory is valid in precisely those cases in which the functional integrals can be evaluated without regard to their rigorous definitions as

[29] If one applies these procedures to the product of two or more operators $F(t''')$, $G(t''')\cdots$ which are functions of the canonical variables taken at different times, then one obtains the matrix elements of the *time ordered* product $[F(t''')G(t''')\cdots]_+$ provided the path summation is understood to be carried out only over paths which contain no parts moving backwards in time.

infinitely multiple integrals (and hence in which the distinction between $\vec{p}_i$ and $\overleftarrow{p}_i$, for example, is unnecessary), one may, in these cases, write purely formally

$$\delta\langle q'',t''|q',t'\rangle = i\hbar^{-1}\int_{q',t'}^{q'',t''}\left\{\delta\int_{t'}^{t''}Ldt\exp\left[i\hbar^{-1}\int_{t'}^{t''}Ldt\right]\right\}\delta[q] = i\hbar^{-1}\left(q'',t''\middle|\delta\int_{t'}^{t''}Ldt\middle|q',t'\right), \quad (7.33)$$

the L in the final expression being the operator Lagrangian.

In the more general case it is possible to write an approximate expression for the well-ordered quantum analog of the classical action function. From (5.66), (7.1), and (7.14), we obtain

$$\mathcal{S}_+(q'',t''|q',t') = S(q'',t''|q',t') - \tfrac{1}{2}i\hbar\ln[D(q'',t''|q',t')/(2\pi i\hbar)^n] + o[\hbar^2(t''-t')^2]. \quad (7.34)$$

The well-ordered operator form follows from this by the method indicated in Sec. 5. In the case of a free particle in one dimension we have exactly

$$\mathcal{S}(q(t''),t''|q(t'),t') = \tfrac{1}{2}(t''-t')^{-1}[q^2(t'')-2q(t'')q(t')+q^2(t')] + \tfrac{1}{2}i\hbar\ln[2\pi i\hbar(t''-t')]. \quad (7.35)$$

The variation of this operator is given by

$$\delta\mathcal{S} = (t''-t')^{-1}[q(t'')-q(t')][\delta q(t'')-\delta q(t')] - \tfrac{1}{2}(t''-t')^{-2}[q^2(t'')-2q(t'')q(t')+q^2(t')](\delta t''-\delta t') + \tfrac{1}{2}i\hbar(t''-t')^{-1}(\delta t''-\delta t'), \quad (7.36)$$

which, since $[q(t''),q(t')] = -i\hbar(t''-t')$, is the same as the variation of the operator

$$S(t''|t') = \tfrac{1}{2}(t''-t')^{-1}[q(t'')-q(t')]^2 = \frac{1}{2}\int_{t'}^{t''}\dot{q}^2(t)dt, \quad (7.37)$$

when written in the form indicated.

Equations of Motion in General Relativity Theory and the Action Principle

LEOPOLD INFELD
Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland

1. HISTORICAL INTRODUCTION

THE history of the problem of motion in general relativity theory usually, and rightly so, starts with the 1927 paper by Einstein and Grommer.[1] This showed for the first time that the equations of motion for a test particle (with mass $m \rightarrow 0$) need not be added to the field equations, but that they can be deduced from the relativistic field equations. For many years afterward, Einstein, and then Einstein with his collaborators, tackled the problem of the motion of two particles. Their problem was to find whether the equations of motion can also be deduced from the field equations. The answer was given in the 1938 paper by Einstein, Infeld, and Hoffmann[2] in which the two-body problem was solved for the first time.

Independently of us and a little later Fock[3] (1939) also deduced the equations of motion, though only the Newtonian ones, from the field equations. Later Papapetrou[4] (1951) simplified his procedure and deduced the post-Newtonian equations of motion, after Petrova[5] (1949) had done the same thing on the basis of Fock's theory. Petrova's and Papapetrou's results were the same as ours.

What are the essential similarities and differences between Einstein's theory, especially as formulated in the two later papers, (Einstein-Infeld, 1940 and 1949)[2] and the Fock-Papapetrou papers?

In general they have one essential idea in common, but two different ideas, of which only one is essential. The idea common to both sets of papers is the approximation method. We now take up the first difference.

In the school represented by Einstein the field equations in empty space are* in the usual notation:

$$G_{\alpha\beta}=R_{\alpha\beta}-\tfrac{1}{2}g_{\alpha\beta}R=0. \tag{1.1}$$

Einstein always thought that to use

$$G_{\alpha\beta}=R_{\alpha\beta}-\tfrac{1}{2}g_{\alpha\beta}=-8\pi T_{\alpha\beta}, \tag{1.2}$$

$T_{\alpha\beta}$ being the energy-momentum tensor, instead of (1.1) is somehow in bad taste, because we do not know in (1.2) what $T_{\alpha\beta}$ is, and we mix a geometrical tensor on the left side with a physical tensor on the right side. This was the reason for Einstein's long search for a unified field theory in which such a mixture of physics and geometry would not appear.

We know that there is no solution of (1.1) representing spread-out continuous matter. Therefore, by assuming (1.1) we represent matter by means of singularities. The method first used by us consisted in forming certain two-dimensional surface integrals over surfaces enclosing these singularities. The field equations prescribed the laws by which the surfaces enclosing the singularities, and hence these singularities, moved. Therefore, these laws were deduced from the field equations in the post-Newtonian approximation.

However, Fock and Papapetrou consider (1.2) and use definite expressions for $T_{\alpha\beta}$.

This difference does not seem to me to be an essential one for the following reasons. We do not know the real distribution of matter. Neither of these methods depicts reality properly. The use of our method based on (1.1) means: if the two bodies are a great distance apart so we may assume, approximately, central symmetry of the field near one body, then the exact knowledge of density distribution inside the enclosing surface is not essential. Outside the enclosing surfaces, (1.1) is valid.

This difference between the two schools may be characterized by an example of a simple situation: that is, classical gravitational theory. There we have two kinds of equations depending on whether matter is represented by singularities or by a continuous distribution. In the first case we have Laplace's equation

$$\Delta\varphi=0; \tag{1.3}$$

in the second case Poisson's equation

$$\Delta\varphi=4\pi\rho. \tag{1.4}$$

It is more common to write Laplace's equation when thinking about its spherically symmetric solution in the form

$$\Delta\varphi=4\pi m\delta_{(3)}, \tag{1.5}$$

where m is the mass and $\delta_{(3)}$ is the three-dimensional Dirac δ function.

Let us try to use the δ functions consistently in general relativity theory. Here we have in the first approximation, using (1.1) for empty space, the

[1] A. Einstein and J. Grommer, Sitzer. deut. Akad. Wiss. Berlin 2 (1927).
[2] Einstein, Infeld, and Hoffman, Ann. Math. 39, 66 (1938). A. Einstein and L. Infeld, Ann. Math. 41, 797 (1940); Can. J. Math. 1, 209 (1949).
[3] V. Fock, J. Phys. (U.S.S.R.) 1, 81 (1939).
[4] A. Papapetrou, Proc. Phys. Soc. (London) 64, 57 (1951).
[5] N. Petrova, J. Phys. (U.S.S.R.) 19, 989 (1949).
* Greek indices run from 0 to 3, Latin from 1 to 3. Repetition implies summation. The quadratic form for a geodetic coordinate-system is

$$ds^2=\eta_{\alpha\beta}dx^\alpha dx^\beta;\quad \eta_{00}=1;\quad \eta_{0m}=0;\quad \eta_{mn}=-\delta_{mn}.$$

The velocity of light $c=1$.

Laplace equation which we solve by a central symmetric solution. This means that a solution of (1.1) in the first approximation valid everywhere, corresponds to that of (1.2) where the energy momentum tensor $T_{\alpha\beta}$ is proportional to Dirac's δ function. Thus the use of (1.2) with $T_{\alpha\beta}$ proportional to $\delta_{(3)}$ corresponds exactly to our previous considerations of (1.1) with singular solutions, just as Eq. (1.5) is only a different form of (1.3).

Use of (1.2) with $T_{\alpha\beta}$ proportional to $\delta_{(3)}$ functions tremendously simplifies the entire deduction of the equations of motion. This simplification was achieved in my paper[6] of 1954, but the entire procedure can still be clarified and simplified. I present here the derivation of the post-Newtonian equations of motion with almost no tedious calculations.†

Summarizing, though I believe in Einstein's idea of not using the energy momentum tensor, yet I was unfaithful to it, because I used the energy-momentum tensor as proportional to the $\delta_{(3)}$ function to express the singular solutions.

The next difference between Fock's and Einstein's school is more essential. Fock, Petrova, and Papapetrou use the harmonic coordinate system; that is the four equations

$$(-g^{\frac{1}{2}}g^{\alpha\beta})_{,\beta}=\frac{\partial}{\partial x^{\beta}}(-g^{\frac{1}{2}}g^{\alpha\beta}). \tag{1.6}$$

Fock considers the choice of this coordinate system to be extremely important, claiming that its addition to the gravitational equations (plus some conditions at infinity) restricts the coordinate system up to a Lorentz transformation. Thus, for Fock, the choice of the harmonic coordinate condition becomes a fundamental law of nature changing the character of Einstein's general relativity theory into a theory of the gravitational field, valid only in inertial coordinate systems. Others like Papapatrou[4] who based their research on Fock's work, do not go so far, but also regard the coordinate condition (1.6) as essential for deduction of the equations of motion.

In our first paper[2] we used a coordinate condition different from the harmonic one; yet we obtained the same equations of motion later obtained by Papapetrou[4] and Petrova.[5] In a few subsequent papers[6–8] we carefully analyzed the problem.

I maintain that the equations of motion have nothing whatever to do with harmonic coordinate conditions; they have much to do with the method of approximation. It is this method which uniquely determines the equations of motion up to the post Newtonian order. I will show explicitly later that violation of the harmonic coordinate condition does not change the post Newtonian equations of motion as long as we stick to the approximation procedure.

2. SOME NOTATIONS AND MATHEMATICAL PRELIMINARIES

We have a world line $\xi^k(t)$ and a field, say a scalar field φ that depends on coordinates x^k and on time $x^0=t$ and also on the $\xi^k(t)$ and their time derivatives:

$$\varphi=\varphi(x^k,t,\xi^k,\dot{\xi}^k);\quad \dot{\xi}^k=d\xi^k/dt. \tag{2.1}$$

We assume that near the line $\xi^k(t)$ the field becomes singular and has the following form:

$$\varphi=\frac{{}^{-1}\varphi}{\rho}+\bar{\varphi}+\overline{\varphi_{,s}}(x^s-\xi^s)+\tfrac{1}{2}\overline{\varphi_{,sr}}(x^s-\xi^s)(x^r-\xi^r)+\cdots. \tag{2.2}$$

Here

$$\rho^2=(x^s-\xi^s)(x^s-\xi^s) \tag{2.3}$$

and, therefore,

$$\bar{\varphi}=\left(\varphi-\frac{{}^{-1}\varphi}{\rho}\right)\bigg|_{x^s=\xi^s}. \tag{2.4}$$

Similarly

$$\overline{\varphi_{,s}}=\frac{\partial}{\partial x^s}\left(\varphi-\frac{{}^{-1}\varphi}{\rho}\right)\bigg|_{x^s=\xi^s}. \tag{2.5}$$

We must distinguish between

$$\overline{\varphi_{,s}}\quad\text{and}\quad \bar{\varphi}_{,s}=\frac{\partial\bar{\varphi}}{\partial\xi^s}=\bar{\varphi}_{,\xi^s}, \tag{2.6}$$

which generally are not equal to each other.

If φ were not singular on the curve $\xi^k(t)$ we could have defined $\bar{\varphi}$ in the following way:

$$\bar{\varphi}=\int\varphi\delta_{(3)}(x^s-\xi^s)d_{(3)}x. \tag{2.7}$$

We can narrow the definition of Dirac's δ functions so that (2.7) remains true even if φ has a singularity up to the kth order. Such δ functions can be constructed (Appendix A) as limits of ordinary functions. By use of such δ functions, we get rid of infinities without recourse to the renormalization procedure. Thus, all δ functions used here will have the property of changing φ into $\bar{\varphi}$, where $\bar{\varphi}$ is a continuous function of the ξ^k, $\dot{\xi}^k$, $\ddot{\xi}^k$. Therefore, we use (2.7) as the definition of $\bar{\varphi}$, where δ is the three-dimensional Dirac δ function, satisfying the following conditions.

1. $\delta(x)$ can be treated formally as a spherically symmetric function for which all the derivatives exist.
2. $\delta(x)=0$ for $x\neq 0$.
3. For every continuous $f(x)$ in the arbitrary region

---

[6] L. Infeld, Phys. Rev. 53, 836 (1938); Can. J. Math. 5, 17 (1953); Acta Phys. Polon. 13, 205 (1954).

† Some of the ideas presented are due to Mr. Plebański and myself and will appear in fuller form in a book *Relativity of Motion* which we are writing.

[7] A. Einstein and L. Infeld, Ann. Math., Can. J. Math. 1, 209 (1949).

[8] R. Teisseyre, Acta Phys. Polon. 13, 47 (1954).

$\Omega(\mathbf{x}_0)$ forming a neighborhood of $\mathbf{x}_0$, we have

$$\int_{\Omega(\mathbf{x}_0)} d_{(3)}x\,\delta(\mathbf{x}-\mathbf{x}_0)f(\mathbf{x})=f(\mathbf{x}_0).$$

4. For an arbitrary neighborhood $\Omega(0)$ of the point $\mathbf{x}=0$, we have

$$\int_{\Omega(0)} d_{(3)}x\,\delta(\mathbf{x})\,|\mathbf{x}|^{-p}=0, \quad \text{for } p=1,2,\cdots k.$$

The fourth condition distinguishes *this* δ function from the usual one. Proof of the inexistence is given in Appendix A. (See also Infeld and Plebański.[9])

Thus the bars mean two things: firstly, singularities are ignored; secondly, for x^k the ξ^k's are introduced.

Returning to the problem of the difference between $\bar{\varphi}_{,s}=\partial\varphi/\partial\xi^s$ and $\overline{\varphi_{,s}}$, we have from (2.7)

$$\bar{\varphi}_{,\xi^s}=\overline{\varphi_{,\xi^s}}+\overline{\varphi_{,s}}. \tag{2.8}$$

This means that $\bar{\varphi}_{,s}$ and $\overline{\varphi_{,s}}$ are equal if and only if $\overline{\varphi_{,\xi^s}}=0$. This is certainly so if the part of φ that gives a contribution to $\bar{\varphi}$ does not depend on ξ^s.

One more formula plays an important role later and follows from the definition (2.7):

$$\bar{\varphi}_{,0}=\frac{d\bar{\varphi}}{dt}=\overline{\varphi_{,0}}+\overline{\varphi_{,s}}\dot{\xi}^s=\overline{\varphi_{,\alpha}}\dot{\xi}^\alpha; \quad (\dot{\xi}^0=1). \tag{2.9}$$

Assume two functions, e.g., φ and ψ:

$$\varphi=\frac{{}^{-1}\varphi}{\rho}+\bar{\varphi}+\overline{\varphi_{,s}}(x^s-\xi^s)+\tfrac{1}{2}\overline{\varphi_{,sr}}(x^s-\xi^s)(x^r-\xi^r)+\cdots,$$
$$\psi=\frac{{}^{-3}\psi}{\rho^3}+\frac{{}^{-1}\psi}{\rho}+\bar{\psi}+\overline{\psi_{,s}}(x^s-\xi^s)+\cdots. \tag{2.10}$$

We have in this case

$$\overline{\varphi\psi}=\bar{\varphi}\bar{\psi}. \tag{2.11}$$

Forming $\varphi\psi$ and ignoring singular expressions and those that vanish for $x^s=\xi^s$, we are left with the following expressions of order zero in $(x^s-\xi^s)$:

$$\bar{\varphi}\bar{\psi}+{}^{-1}\varphi\overline{\varphi_{,s}}\frac{x^s-\xi^s}{\rho}+{}^{-1}\varphi\overline{\psi_{,s}}\frac{x^s-\xi^s}{\rho}$$
$$+{}^{-3}\psi\tfrac{1}{6}\overline{\varphi_{,srp}}\frac{(x^s-\xi^s)(x^r-\xi^r)(x^p-\xi^p)}{\rho^3}.$$

The bars over these expressions give

$$\overline{\varphi\psi}=\int\delta_{(3)}d_{(3)}x\Bigg[\bar{\varphi}\bar{\psi}+(\overline{\varphi_{,s}}\,{}^{-1}\psi+\overline{\psi_{,s}}\,{}^{-1}\varphi)\frac{x^s-\xi^s}{\rho}$$
$$+\tfrac{1}{6}\overline{\varphi_{,srp}}\,{}^{-3}\psi\frac{(x^s-\xi^s)(x^r-\xi^r)(x^p-\xi^p)}{\rho^3}\Bigg].$$

[9] L. Infeld and J. Plebański, Bull. Acad. Polon. III 4, 689 (1956); 5, 51 (1957).

The last two expressions under the integral sign give zero, because they are products of *symmetric* δ functions and *odd* powers of $(x^s-\xi^s)$. Thus we obtain

$$\overline{\varphi\psi}=\int\bar{\varphi}\bar{\psi}\delta d_{(3)}x=\bar{\varphi}\bar{\psi}.$$

But (2.11) would not be true if φ or ψ had a singularity of order ρ^{-2}. Thus (2.11) is true if ρ appears to an *odd* power in the singular parts of φ and ψ. Therefore, we have to be cautious in applying the last equation.

Thus, we have vectors and tensors defined only *along* the curve, like $\bar{S}^\alpha$, $\bar{T}^{\alpha\beta}$, etc. We can define the metric tensor *along* the *curve* $\overline{g_{\alpha\beta}}$ and $\overline{g^{\alpha\beta}}$ and assuming (2.11) we have

$$\overline{g_{\alpha\rho}g^{\beta\rho}}=\overline{g_{\alpha\rho}}\,\overline{g^{\beta\rho}}=\delta_\alpha{}^\beta. \tag{2.12}$$

To such tensors we can apply tensor algebra and tensor analysis but *only* along the curve.

Since

$$\overline{g_{\alpha\beta}}=\int_{\Omega(3)} g_{\alpha\beta}\delta d_{(3)}x \tag{2.13}$$

(where $\Omega_{(3)}$ is a small three-dimensional neighborhood surrounding the singularity), $\overline{g_{\alpha\beta}}$ will be a tensor along the curve if $\delta_{(3)}d_{(3)}x$ is an invariant.

The four-dimensional relativistic Dirac δ function $\delta_{(4)}$, is a scalar density because of the invariant equation

$$\int\delta_{(4)}d_{(4)}x=1. \tag{2.14}$$

Now let us take

$$\delta_{(4)}=\delta_{(4)}(x^\alpha-\xi^\alpha), \tag{2.15}$$

where

$$\xi^\alpha=\xi^\alpha(\lambda), \tag{2.16}$$

λ being an invariant parameter. Then we can form an invariant density function

$$\int_{-\infty}^{+\infty}\delta_{(4)}d\lambda=\int_{-\infty}^{+\infty}\delta_{(4)}\frac{d\lambda}{d\xi^0}d\xi^0=\frac{d\lambda}{dt}\delta_{(3)}. \tag{2.17}$$

As far as the transformation properties are concerned, this is the definitions of $\delta_{(3)}$. Thus $\delta_{(3)}$ is the zero component of a density vector. From this definition it follows that

$$\int_{\Omega(3)}\delta_{(3)}d_{(3)}x=\int_{\Omega(3)}\delta_{(3)}\frac{d\lambda}{dt}d_{(3)}x\frac{dt}{d\lambda}$$
$$=\int_{\Omega(3)}\int_{t=-\infty}^{t=+\infty}\delta_{(4)}d_{(3)}x\,dt=\int_{\Omega(4)}\delta_{(4)}d_{(4)}x. \tag{2.18}$$

Thus the space integral of $\delta_{(3)}$ is an invariant.

One more remark concerning the notation: If we have many curves we shall distinguish between them by the index written above to the left: ${}^a\xi^s$; $a=1,2\cdots p$. Then

we should also write ${}^a\bar{\varphi}$, meaning:

$$ {}^a\bar{\varphi} = \int_{{}^a\Omega_{(3)}} \varphi \, {}^a\delta d_{(3)}x ; \quad {}^a\delta = \delta_{(3)}(x^s - {}^a\xi^s). \quad (2.19) $$

However, for simplicity we do not write a above the bar always understanding that it means the *first* (or only) curve: $\bar{\varphi} = {}^1\bar{\varphi}$.

3. THE GRAVITATIONAL EQUATIONS

The gravitational equations expressed in contravariant tensor densities are

$$ \mathbf{G}^{\alpha\beta} = \mathbf{R}^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta}\mathbf{R} = -8\pi \mathbf{T}^{\alpha\beta}. \quad (3.1) $$

In our case of, say, two particles moving along lines ${}^1\xi^s$ and ${}^2\xi^s$ we have

$$ T^{\alpha\beta} = {}^1t^{\alpha\beta}\, {}^1\delta + {}^2t^{\alpha\beta}\, {}^2\delta. \quad (3.2) $$

We introduce an invariant and finite line element, concerning the first particle (without writing the "one" to the left)

$$ \overline{ds}^2 = \overline{g_{\alpha\beta}} d\xi^\alpha d\xi^\beta . \quad (3.3) $$

From the tensor density we now form a *tensor* along a curve (the first one)

$$ \frac{dt}{\overline{ds}} \int_{\Omega_{(3)}} \mathbf{T}^{\alpha\beta} d_{(3)}x = \overline{t^{\alpha\beta}} \frac{dt}{\overline{ds}} . \quad (3.4) $$

In the next section we see that consistency of the field equations demands the following equations (Tulczyjew[10]):

$$ t^{\alpha\beta} \frac{dt}{\overline{ds}} = \mu \xi^{\alpha\prime} \xi^{\beta\prime} ; \quad \xi^{\alpha\prime} = \frac{d}{\overline{ds}} \xi^\alpha ; \quad \mu = \text{rest mass}. \quad (3.5) $$

We show (Sec. 4) that not only has $t^{\alpha\beta}$ the form (3.5), but also that the rest mass μ is constant. For the moment, however, we use (3.5) as an *assumption*, without stipulating that μ in (3.5) is a constant.

From (3.5) follows

$$ t^{\alpha\beta} = \mu \dot{\xi}^\alpha \xi^{\beta\prime} = m \dot{\xi}^\alpha \dot{\xi}^\beta ; \quad \dot{\xi}^\alpha = \frac{d}{dt}(\xi^\alpha) $$
$$ \frac{m}{\mu} = \frac{dt}{\overline{ds}}, \quad (3.6) $$

and

$$ \mathbf{T}^{\alpha\beta} = \sum_{a=1}^{p} {}^a m \, {}^a\dot{\xi}^\alpha \, {}^a\dot{\xi}^\beta \, {}^a\delta . \quad (3.7) $$

The right-hand side of the gravitational equations is uniquely determined by the condition that $\mathbf{T}^{\alpha\beta}$ depends linearly on the ${}^a\delta$'s.

4. THE GENERAL EQUATIONS OF MOTION

As a consequence of Bianchi's identities we always have

$$ \mathbf{G}^{\alpha\beta}{}_{;\beta} = (\mathbf{R}^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta}\mathbf{R})_{;\beta} = -8\pi \mathbf{T}^{\alpha\beta}{}_{;\beta} = 0 \quad (4.1) $$

where the semicolon indicates covariant differentiation. Taking the integral of $\mathbf{T}^{\alpha\beta}{}_{;\beta}$ over the three-dimensional region surrounding the first singularity and multiplying by $dt/\overline{ds}$, we have:

$$ \frac{dt}{\overline{ds}} \int_{\Omega_{(3)}} \mathbf{T}^{\alpha\beta}{}_{;\beta} d_{(3)}x = \overline{A^\alpha} = 0, \quad (4.2) $$

where $\overline{A^\alpha}$ is a vector defined through (4.2) along the *first* curve. Generally, we have

$$ {}^a\overline{A^\alpha} = 0 ; \quad \alpha = 0, 1, 2, 3 ; \quad a = 1, 2 \cdots p \quad (4.3) $$

if p is the number of singularities. Thus in (4.3) we have as a consequence of the field equations, $4p$ equations which we call the *equations of motion* of the p singularities. Indeed they contain $4p$ unknowns:

$$ {}^a\xi^s(t) \text{ and } {}^a m(t) ; \quad s = 1, 2, 3 ; \quad a = 1, 2 \cdots p. \quad (4.4) $$

Let us now write out $\overline{A^\alpha}$ explicitly (omitting the "one" on the left). We start with

$$ \mathbf{T}^{\alpha\beta}{}_{;\beta} = \mathbf{T}^{\alpha\beta}{}_{,\beta} + \left\{ {\alpha \atop \mu\nu} \right\} \mathbf{T}^{\mu\nu}. \quad (4.5) $$

Introducing here (3.8) the first right-hand expression becomes

$$ \mathbf{T}^{\alpha\beta}{}_{,\beta} = \mathbf{T}^{\alpha s}{}_{,s} + \mathbf{T}^{\alpha 0}{}_{,0} = ({}^1m \, {}^1\dot{\xi}^\alpha \, {}^1\dot{\xi}^s \, {}^1\delta)_{,s} + ({}^1m \, {}^1\dot{\xi}^\alpha \, {}^1\delta)_{,0} \quad (4.6) $$

+similar expressions concerning other particles. We have

$$ \frac{dt}{\overline{ds}} \int_{\Omega_{(3)}} ({}^1m \, {}^1\dot{\xi}^\alpha \, {}^1\dot{\xi}^s \, {}^1\delta)_{,s} d_{(3)}x = \frac{dt}{\overline{ds}} \int_{{}^1\Omega_{(3)}} \mathbf{T}^{\alpha s}{}_{,s} d_{(3)}x = 0. \quad (4.7) $$

This we can see even without explicit calculations by changing the volume integral into a surface integral which must vanish because ${}^1\delta$ vanishes on the surface of ${}^1\Omega_{(3)}$. Therefore, what remains of the integral of (4.6) is, because of (3.6),

$$ \frac{dt}{\overline{ds}} \int_{{}^1\Omega_{(3)}} \mathbf{T}^{\alpha 0}{}_{,0} d_{(3)}x = \frac{dt}{\overline{ds}} \frac{d}{dt} \left[\int_{\Omega_{(3)}} {}^1m \, {}^1\dot{\xi}^\alpha \, {}^1\delta_{(3)} d_{(3)}x \right] $$
$$ = \frac{dt}{\overline{ds}} \frac{d}{dt} (m \dot{\xi}^\alpha) = (\mu \xi^{\alpha\prime})'. \quad (4.8) $$

Thus because of (4.5), (3.6), and the last equation, our equations of motion (4.2) become

$$ \mu' \xi^{\alpha\prime} + \mu \xi^{\alpha\prime\prime} + \mu \overline{\left\{ {\alpha \atop \mu\nu} \right\}} \xi^{\mu\prime} \xi^{\nu\prime} = 0. \quad (4.9) $$

We show now that $\mu' = d\mu/\overline{ds} = 0$, so the rest mass is constant. We assume that

$$ \overline{\left\{ {\alpha \atop \mu\nu} \right\}} = \overline{g^{\alpha\rho}} [\overline{\mu\nu,\rho}]. \quad (4.10) $$

[10] W. Tulczyjew, Bull. Acad. Polon. III, 5, 279 (1957).

Then, multiplying (4.9) by $\bar{g}_{\alpha\sigma}\xi^{\sigma'}$, we have

$$\mu'\xi^{\alpha'}\xi^{\sigma'}\overline{g_{\alpha\sigma}}+\mu\xi^{\alpha''}\xi^{\sigma'}\overline{g_{\alpha\sigma}}+\tfrac{1}{2}\mu\overline{g_{\mu\nu,\sigma}}\xi^{\mu'}\xi^{\nu'}\xi^{\sigma'}=0. \quad (4.11)$$

Because of (2.9) this is equal to

$$\mu'+\frac{\mu}{2}\frac{d}{\overline{ds}}(\overline{g_{\alpha\beta}}\xi^{\alpha'}\xi^{\beta'})=0;\quad \mu'=0, \quad (4.12)$$

which changes (4.9) into

$$\frac{1}{\mu}\overline{A^{\alpha}}=\xi^{\alpha''}+\overline{\left\{ {\alpha \atop \mu\nu} \right\}}\xi^{\mu'}\xi^{\nu'}=0. \quad (4.13)$$

Thus the rest mass must be constant.

If not for the bar above the Christoffel symbol, our Eq. (4.13) would be that of a geodetic line. For each particle their are four such equations of motion.

Let us put back t instead of $\bar{s}$ into (4.13) using (3.6). We then have:

$$\frac{dm}{dt}+m\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}=0,$$
$$\frac{d}{dt}(m\dot\xi^{s})+m\overline{\left\{ {k \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}=0;\quad (\dot\xi^{0}=1). \quad (4.14)$$

As before, we have $4p$ equations determining the ${}^{a}\xi^{k}$ and the ${}^{a}m$'s. From the first equation in (4.14) and (3.8) it follows that

$$\frac{m}{\mu}=\exp\left(-\int^{t}\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}dt\right)=\frac{dt}{\overline{ds}}, \quad (4.15)$$

since

$$\frac{\dot m}{m}=-\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}=\left(\log\frac{m}{\mu}\right)_{,0}. \quad (4.16)$$

From (4.15) we see that

$$\overline{ds}=dt\exp\int_{0}^{t}\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}dt.$$

This connection between $\overline{ds}$ and dt follows from the equations of motion. The normalization is such that for $t=0$, both dt and $\overline{ds}$ are equal.

Substituting the values (4.16) for $\dot m/m$ in the last three equations (4.14) we obtain the $3p$ equations of motion for ${}^{a}\xi^{s}$:

$$\ddot\xi^{s}+\overline{\left\{ {s \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}-\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{s}\dot\xi^{\alpha}\dot\xi^{\beta}=0;\quad \dot\xi^{0}=1. \quad (4.17)$$

This equation is not suitable for establishing a connection between the equations of motion and a variational principle. To do that rewrite the zeroth equation of (4.14) using (3.6)

$$\left(\log\frac{m}{\mu}\right)_{,0}=\left(\log\frac{dt}{\overline{ds}}\right)_{,0}=-\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta} \quad (4.18)$$

or, since $dt/\overline{ds}=(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})^{-\frac{1}{2}}$:

$$[\log(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})^{-\frac{1}{2}}]_{,0}=-\tfrac{1}{2}(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})^{-1}(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})_{,0}$$
$$=-\overline{\left\{ {0 \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}. \quad (4.19)$$

Therefore, the three equations (4.17) can also be written

$$\ddot\xi^{k}+\overline{\left\{ {k \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}-\tfrac{1}{2}(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})^{-1}(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})_{,0}\dot\xi^{k}=0. \quad (4.20)$$

Let us write, for short

$$\overline{\mathfrak{L}}=(\overline{g_{\alpha\beta}}\dot\xi^{\alpha}\dot\xi^{\beta})^{\frac{1}{2}};\quad (\dot\xi^{0}=1). \quad (4.21)$$

Then we can rewrite (4.17) in the form

$$\ddot\xi^{\sigma}-(\log\overline{\mathfrak{L}})_{,0}\dot\xi^{\sigma}+\overline{\left\{ {\sigma \atop \alpha\beta} \right\}}\dot\xi^{\alpha}\dot\xi^{\beta}=0. \quad (4.22)$$

The "zero" equation ($\sigma=0$) gives the known equation (4.19). We multiply (4.22) by $\overline{g_{\sigma r}}$ and assume, that for $\overline{g_{\alpha\beta}}$ and their derivatives (2.11) is always valid, that is the barred product is equal to the products of the barred expressions. Then we have, because of (2.9) the three equations of motion,

$$(\overline{g_{\sigma r}}\dot\xi^{\sigma})_{,0}-\overline{g_{\sigma r}}\dot\xi^{\sigma}(\log\overline{\mathfrak{L}})_{,0}-\tfrac{1}{2}\overline{g_{\alpha\beta,r}}\dot\xi^{\alpha}\dot\xi^{\beta}=0. \quad (4.23)$$

This suggests the existence of a Lagrangian: $\overline{\mathfrak{L}}$. We wish to see whether (4.23) is equivalent to

$$\frac{d}{dt}\frac{\partial\overline{\mathfrak{L}}}{\partial\dot\xi^{k}}-\frac{\partial\overline{\mathfrak{L}}}{\partial\xi^{k}}=0. \quad (4.24)$$

We find

$$\frac{\partial\overline{\mathfrak{L}}}{\partial\xi^{k}}=\tfrac{1}{2}\overline{\mathfrak{L}}^{-1}\overline{g_{\alpha\beta,k}}\dot\xi^{\alpha}\dot\xi^{\beta},$$
$$\frac{\partial\overline{\mathfrak{L}}}{\partial\dot\xi^{k}}=\tfrac{1}{2}\left(\frac{\partial\overline{g_{\alpha\beta}}}{\partial\dot\xi^{k}}\dot\xi^{\alpha}\dot\xi^{\beta}+2\overline{g_{\alpha k}}\dot\xi^{\alpha}\right)\overline{\mathfrak{L}}^{-1}, \quad (4.25)$$
$$\frac{d}{dt}\frac{\partial\overline{\mathfrak{L}}}{\partial\dot\xi^{k}}=\tfrac{1}{2}\left(\frac{\partial\overline{g_{\alpha\beta}}}{\partial\dot\xi^{k}}\dot\xi^{\alpha}\dot\xi^{\beta}\overline{\mathfrak{L}}^{-1}\right)_{,0}+(\overline{g_{\alpha k}}\dot\xi_{\alpha})_{,0}\overline{\mathfrak{L}}^{-1}$$
$$+\overline{g_{\alpha k}}\dot\xi^{\alpha}(\overline{\mathfrak{L}}^{-1})_{,0}.$$

Therefore the Lagrange equation (4.24) becomes

$$(\overline{g_{\alpha k}}\ \dot{\xi}^\alpha)_{,0}-\overline{g_{\alpha k}}\ \dot{\xi}^\alpha(\log\overline{\mathfrak{L}})_{,0}-\tfrac{1}{2}\overline{g_{\alpha\beta,k}}\dot{\xi}^\alpha\dot{\xi}^\beta$$
$$+\frac{1}{2}\left(\frac{\partial\overline{g_{\alpha\beta}}}{\partial\dot{\xi}^k}\dot{\xi}^\alpha\dot{\xi}^\beta\ \overline{\mathfrak{L}}^{-1}\right)_{,0}\overline{\mathfrak{L}}=0. \quad (4.26)$$

Comparing (4.26) and (4.23) we see that $\overline{\mathfrak{L}}$ is a Lagrangian if

$$\overline{g_{\alpha\beta,s}}=\overline{g_{\alpha\beta,s}}\frac{\partial\overline{g_{\alpha\beta}}}{\partial\xi^s},\frac{\partial\overline{g_{\alpha\beta}}}{\partial\dot{\xi}^s}=0. \quad (4.27)$$

This means: $\overline{ds}/dt$ is a Lagrangian if the relevant part of $g_{\alpha\beta}$ does not depend on ${}^1\xi^s$ or ${}^1\dot{\xi}^s$. We shall see later that the first condition, that is the independence of the relevant part of $g_{\alpha\beta}$ from ${}^1\xi^s$ is not always fulfilled.

Thus (4.23) follows from a variational principle:

$$\delta\int_{t_1}^{t_2}\overline{\mathfrak{L}}dt$$
$$=\delta\int_{t_1}^{t_2}dt\left[\int_{\Omega(3)}d_{(3)}x\delta_{(3)}(x^s-\xi^s)g_{\alpha\beta}\dot{\xi}^\alpha\dot{\xi}^\beta\right]^{\frac{1}{2}}=0, \quad (4.28)$$

if we *treat* $g_{\alpha\beta}$ *as function* of x^α *only* that is if, while varying the function under the integration sign, we ignore the possible dependence of $g_{\alpha\beta}$ on ${}^1\xi$ and ${}^1\dot{\xi}$. Then, as can also be shown by a straightforward calculation, (4.28) is equivalent to (4.23). Call the result of such variation the "geodetic line," then (4.23) is the equation of a "geodetic line."[11]

For a *test* particle—that is if $\mu\rightarrow 0$ and $g_{\alpha\beta}$ is not singular and does not depend on ${}^1\xi$ or ${}^1\dot{\xi}$—Eq. (4.23) is that of a geodetic line and the "—" means only the substitution of ${}^1\xi^s$ for x^s.

One can show (Tulczyjew[10]) that the equations of motion follow from the field equations once we assume the linear dependence of $\mathbf{T}^{\alpha\beta}$ from the δ's, that is the form (3.6) for $t_{\alpha\beta}$ and (3.7) for $\mathbf{T}_{\alpha\beta}$ is a consequence of the field equations.

From Bianchi's identities follows:

$$\frac{dt}{\overline{ds}}\int_{{}^1\Omega(3)}\theta\mathbf{T}^{\alpha\beta}{}_{;\beta}d_{(3)}x=0, \quad (4.29)$$

where θ is an *arbitrary* function continuous on the worldline $\xi^s(t)$. Or, again omitting the "one" over the $t^{\alpha\beta}$'s and δ's we can write

$$\frac{dt}{\overline{ds}}\int_{{}^1\Omega(3)}\theta(t^{\alpha\beta}\delta_{(3)})_{;\beta}d_{(3)}x=0. \quad (4.30)$$

[11] L. Infeld and J. Plebański, Bull. Acad. Polon. III, 4, 749 (1956).

Thus this leads to the following:

$$\overline{A^\alpha}\,\overline{\theta}+\overline{A^{\alpha\beta}}\,\overline{\theta_{,\beta}}=0. \quad (4.31)$$

Let us start by calculating $\overline{A^{\alpha\beta}}$. Because

$$\delta_{,0}=-\delta_{,s}\dot{\xi}^s, \quad (4.32)$$

we have

$$\overline{\theta_{,\beta}}\ \overline{A^{\alpha\beta}}=\frac{dt}{\overline{ds}}(-t^{\alpha s}+t^{\alpha 0}\dot{\xi}^s)\overline{\theta_{,s}}. \quad (4.33)$$

Since $\dot{\xi}^0=1$, this can be written

$$\overline{\theta_{,\beta}}\ \overline{A^{\alpha\beta}}=\frac{dt}{\overline{ds}}(-t^{\alpha\beta}+t^{\alpha 0}\dot{\xi}^\beta)\overline{\theta_{,\beta}}. \quad (4.34)$$

Because $\overline{\theta_{,\beta}}$ is arbitrary, we have

$$\overline{A^{\alpha\beta}}=\frac{dt}{\overline{ds}}(-t^{\alpha\beta}+t^{\alpha 0}\dot{\xi}^\beta)=0. \quad (4.35)$$

Putting here $\alpha=0$ we find

$$t^{0\beta}=t^{00}\dot{\xi}^\beta, \quad (4.36)$$

therefore, generally

$$t^{\alpha\beta}=t^{00}\dot{\xi}^\alpha\dot{\xi}^\beta=m\dot{\xi}^\alpha\dot{\xi}^\beta, \quad (4.37)$$

which is the proof of the theorem. Obviously $\overline{A}^\alpha=0$ gives the equations of motion.

5. THE APPROXIMATION METHOD

We solve the field equations and formulate the equations of motion explicitly by means of the approximation method to be described.

Let us assume a function developed in a power series in the parameter $\lambda=(1/c)$ (assume c arbitrary and not equal to "one").

$$\varphi={}_0\varphi+{}_1\varphi+{}_2\varphi+\cdots. \quad (5.1)$$

The indices written as left subscripts indicate the order of λ absorbed by the φ's.

If the function $\varphi(x^\mu)$ varies rapidly in space but slowly with x^0, then we are justified in not treating all its derivatives in the same manner. The derivatives with respect to x^0 will be of a higher order than the space derivatives. We can formalize this procedure by assuming

$$\frac{\partial}{\partial x^0}({}_l\varphi)={}_{l+1}\varphi_{,0}; \quad (5.2)$$

that is differentiation with respect to x^0 raises the power of λ, absorbed by the φ's, by one.

The problem now is: with what order should we start the power development of the quantities appearing in the field equations?

The quantity ξ^s will start (by an obvious convention) with the order "zero." It will be an unknown quantity

determined by the equations of motion; we shall for the moment not develop ξ^s into a power series. Thus $\dot{\xi}^s$ will be of the order "one" and $\ddot{\xi}^s$ of the order "two."

From this follows

$$ {}^am = {}_2{}^am + {}_4{}^am + {}_6{}^am + \cdots . \tag{5.3} $$

To begin with, ${}_2m$ is not pure convention. Indeed in the Newtonian approximation, which we hope to obtain, we have in the chosen units

$$ \text{mass}\times\text{acceleration} = \frac{\text{mass}\times\text{mass}}{(\text{distance})^2}. $$

Since the acceleration is of order two, the orders of both sides will be equal, only if the order of mass is also "two."

In all the power developments we take into account only even (as in ${}^am$), or only odd powers of λ (Infeld,[6] 1938).

Thus, because of the order with which we start ${}^am$ and ξ^s, we have

$$ \begin{aligned} \mathbf{T}^{00} &= {}_2\mathbf{T}^{00} + {}_4\mathbf{T}^{00} + {}_6\mathbf{T}^{00} + \cdots, \\ \mathbf{T}^{0m} &= {}_3\mathbf{T}^{0m} + {}_5\mathbf{T}^{0m} + \cdots, \\ \mathbf{T}^{mn} &= {}_4\mathbf{T}^{mn} + {}_6T^{mn} + \cdots . \end{aligned} \tag{5.4} $$

Now as to the g's we write

$$ g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}; \quad g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu}. \tag{5.5} $$

From the gravitational equation follows:

$$ \mathbf{R}_{\alpha\beta} = -8\pi(\mathbf{T}_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}\mathbf{T}), \tag{5.6} $$

where

$$ \mathbf{R}_{\alpha\beta} = R_{\alpha\beta}\sqrt{-g}, $$

$$ R_{\mu\nu} = -\left\{ {\rho \atop \mu\nu} \right\}_{,\rho} + \left\{ {\rho \atop \mu\rho} \right\}_{,\nu} + \left\{ {\rho \atop \mu\sigma} \right\}\left\{ {\sigma \atop \rho\nu} \right\} - \left\{ {\rho \atop \mu\nu} \right\}\left\{ {\sigma \atop \rho\sigma} \right\}. \tag{5.7} $$

From the right-hand side of Eq. (5.6) follows: R_{00} and R_{mn} start with the order "2" and R_{0m} with the order "3." The lowest order expressions on the left-hand side are

$$ \begin{aligned} &\text{for } R_{00}: \quad -\tfrac{1}{2}h_{00,ss}, \\ &\text{for } R_{mn}: \quad -\tfrac{1}{2}h_{mn,ss} + \tfrac{1}{2}h_{ms,ns} + \tfrac{1}{2}h_{ns,ms} + \tfrac{1}{2}h_{00,mn} - \tfrac{1}{2}h_{ss,mn}, \quad (5.8) \\ &\text{for } R_{0m}: \quad -\tfrac{1}{2}h_{0m,ss} + \tfrac{1}{2}h_{0s,ms} + \tfrac{1}{2}h_{ms,0s} - \tfrac{1}{2}h_{ss,m0}. \end{aligned} $$

Therefore,

$$ \begin{aligned} h_{00} &= {}_2h_{00} + {}_4h_{00} + \cdots, \\ h_{0m} &= {}_3h_{0m} + {}_5h_{0m} + \cdots, \\ h_{mn} &= {}_2h_{mn} + {}_4h_{mn} + \cdots . \end{aligned} \tag{5.9} $$

All the functions that appear later are obtained from the h's by summation, multiplication, differentiation. To every component, the following rule applies: any component having an odd (even) number of zero suffixes has only odd (even) powers of λ in its expansion.

6. THE NEWTONIAN EQUATIONS OF MOTION

We try to find the equations of motion in the lowest (Newtonian) approximation. We do it in such a way as to make the generalization to the post-Newtonian approximation as simple as possible.

Because of (5.6) and (5.8), the field equations of the lowest order are in h_{00}

$$ -\tfrac{1}{2}\,{}_2h_{00,ss} = -8\pi({}_2\mathbf{T}^{00} - \tfrac{1}{2}\,{}_2\mathbf{T}^{00}) = -4\pi\,{}_2\mathbf{T}^{00} = -4\pi \sum_{a=1}^{p} {}_2{}^am\,{}^a\delta, \tag{6.1} $$

or

$$ {}_2h_{00,ss} = 8\pi \sum_{a=1}^{p} {}_2{}^am\,{}^a\delta. \tag{6.2} $$

As the solution of this equation we take the Newtonian field, that is

$$ {}_2h_{00} = -2\,{}_2{}^1m\,{}^1r^{-1} - 2\,{}_2{}^2m\,{}^2r^{-1} - \cdots, \tag{6.3} $$

where

$$ {}^ar^2 = (x^s - {}^a\xi^s)(x^s - {}^a\xi^s). \tag{6.4} $$

We also write briefly:

$$ {}_2h_{00} = \varphi \tag{6.5} $$

and in the two-body case, which we assume for the sake of simplicity‡

$$ \begin{aligned} &\varphi = f + g, \\ &f = -2\,{}^1m\,{}^1r^{-1}; \quad g = -2\,{}^2m\,{}^2r^{-1}, \\ &{}^1\xi = \xi; \quad {}^2\xi = \eta. \end{aligned} \tag{6.6} $$

Because of (5.5) and

$$ g^{\alpha\rho}g_{\beta\rho} = \delta^\alpha{}_\beta, \tag{6.7} $$

we have

$$ {}_2h^{00} = -\varphi. \tag{6.8} $$

The relevant part of φ, that is g, does not depend on ξ; therefore, we have

$$ \overline{{}_2h_{00,s}} = {}_2\bar{h}_{00,s} = \overline{g_{,s}} = \frac{\partial\bar{g}}{\partial\xi^s} = \bar{g}_{,\xi^s}. \tag{6.9} $$

There is no reason to assume that ${}_2m$ is a constant. This follows, however, from (4.15), or, the first equation in (4.14).

Because of (6.9) and because in the lowest order we do not have any products of the h's, the conditions (4.10) and (4.27) are satisfied, this means that the path is a "geodetic line" and can be deduced from a Lagrangian:

$$ \overline{\mathfrak{L}} = (\overline{g_{\alpha\beta}}\,\dot{\xi}^\alpha\dot{\xi}^\beta)^{\frac{1}{2}}; \quad {}_2\overline{\mathfrak{L}} = (1 - \dot{\xi}^s\dot{\xi}^s + \bar{g})^{\frac{1}{2}}, \tag{6.10} $$

$$ \begin{aligned} {}_2\overline{\mathfrak{L}} &= -\tfrac{1}{2}\dot{\xi}^s\dot{\xi}^s + \tfrac{1}{2}\bar{g} = -\tfrac{1}{2}\dot{\xi}^s\dot{\xi}^s - \frac{{}^2m}{r}, \\ r^2 &= (\xi^s - \eta^s)(\xi^s - \eta^s). \end{aligned} \tag{6.11} $$

‡ There is, I hope, no danger in confusing this g in (6.6) with $g = |g_{\alpha\beta}|$.

Thus the equations of motion, up to the second order, are

$$\frac{d}{dt}\frac{\partial \overline{\mathfrak{L}}}{\partial \dot{\xi}^s}-\frac{\partial \overline{\mathfrak{L}}}{\partial \xi^s}=0, \tag{6.12}$$

that is

$$\ddot{\xi}^s=\frac{\partial}{\partial \xi^s}({}^2m/r). \tag{6.13}$$

The Lagrangian $\overline{\mathfrak{L}^{**}}$ for *both* particles is

$${}_4\overline{\mathfrak{L}^{**}}=-\tfrac{1}{2}\,{}^1m\dot{\xi}^s\dot{\xi}^s-\tfrac{1}{2}\,{}^2m\dot{\eta}^s\dot{\eta}^s-\frac{{}^1m\,{}^2m}{r}. \tag{6.14}$$

The Newtonian Lagrangian for both particles is of the fourth order. Since $(m/\mu)=(\overline{dt/ds})$, we have here

$${}_2m={}_2\mu,\quad {}_4{}^1m=\tfrac{1}{2}\,{}_2{}^1m\dot{\xi}^s\dot{\xi}^s+{}^1m\,{}^2m/r. \tag{6.15}$$

We also find the equations of motion directly by going back to (4.17). We then have simply

$$\ddot{\xi}^s+\overline{\left\{\begin{matrix} s \\ 00 \end{matrix}\right\}}=0 \tag{6.16}$$

or

$$\ddot{\xi}^s-\overline{[00,s]}=\ddot{\xi}^s+\tfrac{1}{2}\overline{g_{,s}}=0, \tag{6.17}$$

which is identical with (6.13).

7. TRANSITION TO THE NEXT APPROXIMATION

To find the equations of motion up to the fourth order, besides ${}_2h_{00}$, we must also know

$${}_2h_{mn},\ {}_3h_{0m},\ {}_4h_{00}. \tag{7.1}$$

The first two are easy. The left-hand side of the corresponding equations is written out in (5.8) and the right-hand side is given by (5.6) and it is

$$\begin{aligned} &\text{for } mn: && -4\pi\delta_{mn}({}_2{}^1m\,{}^1\delta+{}_2{}^2m\,{}^2\delta),\\ &\text{for } 0m: && 8\pi({}_2{}^1m\dot{\xi}^m\,{}^1\delta+{}_2{}^2m\dot{\eta}^m\,{}^2\delta). \end{aligned} \tag{7.2}$$

Therefore, for ${}_2h_{mn}$ we have the equations

$$-\tfrac{1}{2}\,{}_2h_{mn,ss}+\tfrac{1}{2}\,{}_2h_{ms,ns}+\tfrac{1}{2}\,{}_2h_{ns,ms}-\tfrac{1}{2}\,{}_2h_{ss,mn}+\tfrac{1}{2}\varphi_{,mn}=-\tfrac{1}{2}\delta_{mn}\varphi_{,ss}. \tag{7.3}$$

We are looking for a solution of a Newtonian character. Such a solution is

$${}_2h_{mn}=\delta_{mn}\varphi. \tag{7.4}$$

The choice of these Newtonian solutions for ${}_2h_{00}$ and ${}_2h_{mn}$ must be regarded as part of our approximation procedure (Sec. 9).

The next step is to calculate ${}_3h_{0m}$. Again (5.8) and (7.2) gives

$$-\tfrac{1}{2}\,{}_3h_{0n,ss}+\tfrac{1}{2}\,{}_3h_{0s,ns}+\tfrac{1}{2}\,{}_3h_{ns,0s}-\tfrac{1}{2}\,{}_3h_{ss,n0}=8\pi({}_2{}^1m\dot{\xi}^n\delta^1+{}_2{}^2m\dot{\eta}^n\delta^2). \tag{7.5}$$

If (7.4) is introduced into the last equation we have

$$-\tfrac{1}{2}\,{}_3h_{0n,ss}+\tfrac{1}{2}\,{}_3h_{0s,ns}-\varphi_{,n0}=8\pi({}_2{}^1m\dot{\xi}^n\,{}^1\delta+{}_2m^2\dot{\eta}^{n2}\delta). \tag{7.6}$$

The solution which we adopt here is

$${}_3h_{0n}=-2f\dot{\xi}^n-2g\dot{\eta}^n. \tag{7.7}$$

This is not the only possible solution. Section 9 considers a more general solution and its influence upon the equations of motion.

Calculation of ${}_4h_{00}$ is more troublesome (Appendix B). Here we quote the relevant expressions of ${}_4h_{00}$, that is those that give a contribution to ${}_4\overline{h_{00}}$ and ${}_4\overline{h_{00,\alpha}}$. They are

$${}_4h_{00}\sim 2({}^2m)^2({}^2r)^{-2}-3\dot{\eta}^s\dot{\eta}^s\,{}^2m({}^2r)^{-1}-{}^2m\,{}^2r_{,00}+2\,{}^1m\,{}^2m(r\,{}^2r)^{-1}. \tag{7.8}$$

Now we should like to see whether these expressions for ${}_2h_{mn}$, ${}_3h_{0n}$, ${}_4h_{00}$ are such that they make the equations of motion equivalent to those of a "geodetic line," that is, if the conditions

$$\overline{\left\{\begin{matrix} \alpha \\ \beta\rho \end{matrix}\right\}}=\overline{g^{\alpha\sigma}[\beta\rho,\sigma]}, \tag{7.9}$$

$$\overline{g_{\alpha\beta,s}}=\overline{g_{\alpha\beta}}_{,s};\quad \partial\overline{g_{\alpha\beta}}/\partial\dot{\xi}^r=0 \tag{7.10}$$

are satisfied. Both conditions are satisfied for ${}_2h_{mn}$ and ${}_3h_{0m}$. These expressions have a singularity of order $1/r$, an odd singularity; therefore, for them, (7.9) is satisfied. In their relevant part neither ξ nor $\dot{\xi}$ appears explicitly; therefore, for them, (7.10) is satisfied. But this is not true of ${}_4h_{00}$. It has an even singularity because of the appearance of $({}^1r)^{-2}$. Yet this does not matter, since ${}_4h_{00}$ appears only *linearly* in the equations of motion. Therefore, we can disregard the condition (7.9) for ${}_4h_{00}$. But is the condition (7.10) satisfied for ${}_4h_{00}$? It is certainly satisfied for the first two expressions, that is, for $2({}^2m)^2({}^2r)^{-2}-3\dot{\eta}^s\dot{\eta}^s\,{}^2m({}^2r)^{-1}$, since neither of these two expressions depends on $\dot{\xi}$, ξ. We must be more careful with the third expression:

$$-{}^2m\,{}^2r_{,00}=-{}^2m\,{}^2r_{,\eta^s\eta^r}\dot{\eta}^s\dot{\eta}^r-{}^2m\,{}^2r_{,\eta^s}\ddot{\eta}^s. \tag{7.11}$$

The first expression on the right-hand side does not depend on ξ, $\dot{\xi}$, but the second expression contains $\ddot{\eta}^s$. Here we may introduce for ${}^2m\ddot{\eta}^s=-{}^1m\ddot{\xi}^s$, the Newtonian value, since the mistake will be of order "6". Thus we can put

$$\alpha=-{}^2m\,{}^2r_{,\eta^s}\ddot{\eta}^s={}^1m\,{}^2m\,{}^2r_{,\eta^s}\left(\frac{1}{r}\right)_{,\xi^s}$$

$$=[{}^1m\,{}^2m(x^s-\eta^s)({}^2r)^{-1}](\xi^s-\eta^s)r^{-3}. \tag{7.12}$$

This expression, differentiated with respect to $\dot{x}^m$

gives zero at the point $x^s = \xi^s$. Thus

$$\overline{\alpha_{,m}} = 0. \tag{7.13}$$

But

$$\bar{\alpha}_{,m} = \frac{\partial \bar{\alpha}}{\partial \xi^m}\, {}^1m\, {}^2m (r^{-2})_{,\xi^m}.$$

A similar situation occurs in the last expression in ${}_4h_{00}$:

$$\beta = 2\, {}^1m\, {}^2m (r\, {}^2r)^{-1}. \tag{7.14}$$

We have

$$\overline{\beta_{,s}} = 2\, {}^1m\, {}^2m r^{-1} \overline{\left(\frac{1}{r^2}\right)_{,s}} = 2\, {}^1m\, {}^2m r^{-1} (r^{-1})_{,\xi^s},$$

$$\bar{\beta}_{,s} = 2\, {}^1m\, {}^2m (r^{-2})_{,\xi^s}. \tag{7.15}$$

Therefore

$$\bar{\beta}_{,s} = 2\overline{\beta_{,s}}. \tag{7.16}$$

However, we can easily find an auxiliary field which we shall denote by $h_{00}{}^*$ such that

$${}_4\bar{h}^*{}_{00,m} = \overline{{}_4h_{00,m}}. \tag{7.17}$$

From this it follows that such $\overline{{}_4h_{00}{}^*}$ is

$$\overline{{}_4h^*{}_{00}} = 2({}^2m)^2 r^{-2} - 3\dot{\eta}^s\dot{\eta}^s\, {}^2m r^{-1}$$
$$- {}^2m r_{,\xi^s\xi^r}\dot{\eta}^s\dot{\eta}^r + {}^1m\, {}^2m r^{-2}. \tag{7.18}$$

8. THE POST NEWTONIAN EQUATIONS OF MOTION

In the general equations of motion (4.17) ${}_4h_{00}$ will appear only *once* up to the fourth order, that is in the expression $\frac{1}{2}\, \overline{{}_4h_{00,m}}$. This means, because of (7.17), that a Lagrangian up to the fourth order exists§:

$$\overline{\mathfrak{L}^*} = \overline{ds^*}/dt, \tag{8.1}$$

with

$$\left(\frac{\overline{ds^*}}{dt}\right)^2 = \overline{g_{\alpha\beta}{}^*}\, \xi^\alpha \xi^\beta \tag{8.2}$$

in which only $\overline{g_{00}{}^*} \neq \overline{g_{00}}$ and

$$g_{00}{}^* = \eta_{00} + {}_2h_{00} + {}_4h_{00}{}^* = 1 + \varphi + {}_4h_{00}{}^*. \tag{8.3}$$

Therefore,

$$\overline{\mathfrak{L}^*} = (1 + \bar{\varphi} + \overline{{}_4h^*{}_{00}} - \dot{\xi}^s\dot{\xi}^s + \bar{\varphi}\, \dot{\xi}^s\dot{\xi}^s + \overline{{}_3h_{s0}}\, \dot{\xi}^s)^{\frac{1}{2}}. \tag{8.4}$$

Here we introduce the values

$$\bar{\varphi} = -2\, {}^2m r^{-1}; \quad \overline{{}_3h_{0s}} = 4\, {}^2m r^{-1}\dot{\eta}^s;$$
$$\overline{{}_4h_{00}{}^*} = 2({}^2m)^2 r^{-2} - 3\dot{\eta}^s\dot{\eta}^s\, {}^2m r^{-1}$$
$$- {}^2m\, r_{,\xi^s\xi^r}\dot{\eta}^s\dot{\eta}^r + {}^1m\, {}^2m r^{-1}. \tag{8.5}$$

Remembering that up to the fourth order

$$(1 + {}_2a + {}_4a)^{\frac{1}{2}} = 1 + \tfrac{1}{2}({}_2a + {}_4a) - \tfrac{1}{8}\, {}_2a^2 \tag{8.6}$$

we obtain from (8.4) and (8.5) the Lagrangian:

$$\overline{\mathfrak{L}^*} = \overline{{}_2\mathfrak{L}^*} + \overline{{}_4\mathfrak{L}^*} = -\tfrac{1}{2}\dot{\xi}^s\dot{\xi}^s - {}^2m r^{-1}$$
$$- \tfrac{3}{2}\, {}^2m r^{-1}(\dot{\xi}^s\dot{\xi}^s + \dot{\eta}^s\dot{\eta}^s) + 4\, {}^2m r^{-1}\dot{\xi}^s\dot{\eta}^s - \tfrac{1}{8}(\dot{\xi}^s\dot{\xi}^s)^2$$
$$+ \frac{1}{2}\frac{{}^2m({}^1m + {}^2m)}{r^2} - \tfrac{1}{2}\, {}^2m r_{,\xi^s\xi^r}\dot{\eta}^s\dot{\eta}^r. \tag{8.7}$$

This is supposed to be the Lagrangian for the *first* particle. But we wish to find the Lagrangian for *both* particles. This means *first* a Lagrangian which gives the same equations of motion as the Lagrangian (8.7); *secondly*, a Lagrangian which is invariant with respect to a transformation, changing

$${}^1m\xi \rightleftarrows {}^2m\eta. \tag{8.8}$$

Let us multiply $\overline{\mathfrak{L}^*}$ in (8.7) by ${}^1m$. The equations of motion will be the same. Let us add

$$-\tfrac{1}{2}\, {}^2m\dot{\eta}^s\dot{\eta}^s - \tfrac{1}{8}\, {}^2m(\dot{\eta}^s\dot{\eta}^s)^2. \tag{8.9}$$

With this addition, the equations of motion for the first particle will still be the same, since there is no contribution from (8.9). With these changes, the only expression not invariant with respect to the change (8.8) will be the last one in (8.7), that is

$$-\tfrac{1}{2}\, {}^1m\, {}^2m r_{,\xi^s\xi^r}\dot{\eta}^s\dot{\eta}^r, \tag{8.10}$$

But instead of it we can write

$$\tfrac{1}{2}\, {}^1m\, {}^2m r_{,\xi^s\eta^r}\dot{\xi}^s\dot{\eta}^r, \tag{8.11}$$

which is invariant with respect to the change (8.8), and gives *the same* contributions to the equations of motion.

From (8.10) we have the following addition to the equations of motion of the first particle:

$$\tfrac{1}{2}\, {}^1m\, {}^2m r_{,\xi^s\xi^r\xi^n}\dot{\eta}^s\dot{\eta}^r = -\tfrac{1}{2}\, {}^1m\, {}^2m r_{,\xi^s\eta^r\xi^n}\dot{\eta}^s\dot{\eta}^r. \tag{8.12}$$

From (8.11) we have the addition:

$$\tfrac{1}{2}\, {}^1m\, {}^2m (r_{,\xi^n\eta^r}\dot{\eta}^r)_{,0} - \tfrac{1}{2}\, {}^1m\, {}^2m r_{,\xi^s\eta^r\xi^n}\dot{\xi}^s\dot{\eta}^r$$
$$= \tfrac{1}{2}\, {}^1m\, {}^2m (r_{,\xi^n\eta^r\xi^s}\dot{\eta}^r\dot{\xi}^s + r_{,\xi^n\eta^r\eta^s}\dot{\eta}^r\dot{\eta}^s$$
$$+ r_{,\xi^n\eta^r}\ddot{\eta}^r - r_{,\xi^s\eta^r\xi^n}\dot{\xi}^s\dot{\eta}^r)$$
$$= \tfrac{1}{2}\, {}^1m\, {}^2m (r_{,\xi^n\eta^r\eta^s}\dot{\eta}^r\dot{\eta}^s + r_{,\xi^n\eta^r}\ddot{\eta}^r). \tag{8.13}$$

§ Compare I. G. Fichtenholz [J. Phys. (U.S.S.R.) 27, 563 (1954)] where the Lagrangian is found mechanically from the explicit equations of motion and not, as here, the equation of motion from the Lagrangian.

But the last expression in (8.13) equals

$$\tfrac{1}{2}\,{}^1m\,{}^2m r_{,\xi^n\xi^r}\ddot{\eta}^r$$
$$=\tfrac{1}{2}\,{}^1m\,{}^2m\left(-\frac{\delta_{nr}}{r}+\frac{(\xi^n-\eta^n)(\xi^r-\eta^r)}{r^3}\right)\ddot{\eta}^r. \quad (8.14)$$

Since

$$\ddot{\eta}^r=\frac{{}^1m(\xi^r-\eta^r)}{r^3}, \quad (8.15)$$

we see that (8.14) vanishes and, therefore, (8.10) and (8.11) give the same contributions to the equations of motion.

Therefore, if we call the final Lagrangian for two particles $\overline{\mathfrak{L}^{**}}$, we have

$$\overline{\mathfrak{L}^{**}}=\overline{{}_4\mathfrak{L}^{**}}+\overline{{}_6\mathfrak{L}^{**}}=-\tfrac{1}{2}\,{}^1m\dot{\xi}^s\dot{\xi}^s-\tfrac{1}{2}\,{}^2m\dot{\eta}^s\dot{\eta}^s-\frac{{}^1m\,{}^2m}{r}$$
$$-\frac{3}{2}\frac{{}^1m\,{}^2m}{r}(\dot{\xi}^s\dot{\xi}^s+\dot{\eta}^s\dot{\eta}^s)+4\frac{{}^1m\,{}^2m}{r}\dot{\xi}^s\dot{\eta}^s$$
$$-\tfrac{1}{8}\,{}^1m(\dot{\xi}^s\dot{\xi}^s)^2-\tfrac{1}{8}\,{}^2m(\dot{\eta}^s\dot{\eta}^s)^2$$
$$+\frac{1}{2}\frac{{}^1m\,{}^2m({}^1m+{}^2m)}{r^2}+\tfrac{1}{2}\,{}^1m\,{}^2m r_{,\xi^s\eta^r}\dot{\xi}^s\dot{\eta}^r. \quad (8.16)$$

From this Lagrangian the equations of motion for the first particle are:

$$\ddot{\xi}^n-{}^2m\left(\frac{1}{r}\right)_{,\xi^n}$$
$$={}^2m\left\{\left[\dot{\xi}^s\dot{\xi}^s+\tfrac{3}{2}\dot{\eta}^s\dot{\eta}^s-4\dot{\xi}^s\dot{\eta}^s-4\frac{{}^2m}{r}-5\frac{{}^1m}{r}\right]\left(\frac{1}{r}\right)_{,\xi^n}\right.$$
$$+\left[4\dot{\xi}^s(\dot{\eta}^n-\dot{\xi}^n)+3\dot{\xi}^n\dot{\eta}^s-4\dot{\eta}^n\dot{\eta}^s\right]\left(\frac{1}{r}\right)_{,\xi^s}$$
$$\left.+\tfrac{1}{2}r_{,\xi^s\xi^r\xi^n}\dot{\eta}^s\dot{\eta}^r\right\}. \quad (8.17)$$

The equations of motion for the other particle are obtained by replacing

$$ {}^1m\,{}^2m,\ \xi,\ \eta \quad \text{by} \quad {}^2m,\ {}^1m,\ \eta,\ \xi, \quad (8.18)$$

respectively.

The generalization of this result to p particles is almost trivial, if we take into account the changes caused by the addition of these particles in $\overline{h_{00}}^*$ (Appendix B). These additional expressions are due, say, in the case of three particles, to interaction between the second and third particles; that is, in the equations of motion for the first particle they will give a contribution proportional to ${}^1m\,{}^2m\,{}^3m$. These expressions appear in $\overline{\mathfrak{L}^{**}}$ from two sources: from $\overline{{}_4h_{00}}^*$ and from $\overline{\varphi^2}$ in (8.6). If we now denote the "distance" from the ath to the bth particle by:

$$({}^{(ab)}r)^2=({}^a\xi^s-{}^b\xi^s)({}^a\xi^s-{}^b\xi^s), \quad (8.19)$$

then we have the Lagrangian for the p particles:

$$\overline{\mathfrak{L}^{**}}=\overline{{}_4\mathfrak{L}^{**}}+\overline{{}_6\mathfrak{L}^{**}}$$
$$=-\tfrac{1}{2}\sum_{a=1}^{p}{}^am\,{}^a\dot{\xi}^s\,{}^a\dot{\xi}^s-\tfrac{1}{2}\sum_{\substack{a,b=1\\a\neq b}}^{p}{}'\,{}^am\,{}^bm({}^{(ab)}r)^{-1}$$
$$-\tfrac{3}{4}\sum_{\substack{a,b=1\\a\neq b}}^{p}{}'\,{}^am\,{}^bm({}^{(ab)}r)^{-1}({}^a\dot{\xi}^s\,{}^a\dot{\xi}^s+{}^b\dot{\xi}^s\,{}^b\dot{\xi}^s)$$
$$+2\sum_{\substack{a,b=1\\a\neq b}}^{p}{}'\,{}^am\,{}^bm({}^{(ab)}r)^{-1}\,{}^a\dot{\xi}^s\,{}^b\dot{\xi}^s$$
$$-\tfrac{1}{8}\sum_{a=1}^{p}{}^am({}^a\dot{\xi}^s\,{}^a\dot{\xi}^s)^2$$
$$+\tfrac{1}{4}\sum_{\substack{a,b=1\\a\neq b}}^{p}{}'\,{}^am\,{}^bm({}^am+{}^bm)({}^{(ab)}r)^{-2}$$
$$+\sum_{\substack{a,b=1\\a\neq b}}^{p}{}'\,{}^am\,{}^bm({}^{(ab)}r_{,\xi^s\xi^r})\,{}^a\dot{\xi}^s\,{}^b\dot{\xi}^r$$
$$+\tfrac{1}{6}\sum_{\substack{a,b,c=1\\a\neq b\neq c}}^{p}{}''\,{}^am\,{}^bm\,{}^cm\left[({}^{(ab)}r\,{}^{(ac)}r)^{-1}\right.$$
$$\left.+({}^{(bc)}r\,{}^{(ba)}r)^{-1}+({}^{(ca)}r\,{}^{(cb)}r)^{-1}\right]. \quad (8.20)$$

In the case of two particles the Lagrangian (8.20) reduces to (8.16). The only new expression appearing in (8.20) is the last one; in the case of three particles it is equal to

$${}^1m(\tfrac{1}{2}\,\overline{{}_4s_{00}}-\tfrac{1}{4}\bar{g}\bar{k})$$

using the notation of Appendix B where $\overline{{}_4s_{00}}$ is the change in $\overline{{}_4h_{00}}^*$ caused by the interaction of the second and third particles and k is for the third particle what f and g are for the first and second. Thus $-\tfrac{1}{4}\bar{g}\bar{k}$ is the contribution to the Lagrangian of the interaction between the second and third particles coming from $-\tfrac{1}{8}\overline{\varphi^2}$ in (8.6).

9. ON THE CHOICE OF THE COORDINATE SYSTEM

The harmonic coordinate condition is

$$[(-g)^{\frac{1}{2}}g^{\mu\nu}]_{,\nu}=0, \quad (9.1)$$

which in our case means

$${}_2h^{mn}{}_{,n}=0;\quad {}_3h^{0m}{}_{,m}+{}_2h^{00}{}_{,0}=0. \quad (9.2)$$

None of these conditions are fulfilled in our coordinate system.

The values for ${}_2h_{00}$ and ${}_2h_{mn}$ accepted by us here were the Newtonian values. Our convention is that their choice characterized our approximation procedure. Yet,

with some justification, this approach may be regarded as too formal. Instead of our values for ${}_2h_{mn}$ we could have chosen

$${}_2h_{mn}' = {}_2h_{mn} + {}_2a_{m,n} + {}_2a_{n,m}$$

the a's being arbitrary functions. This change from ${}_2h_{mn}$ to ${}_2h_{mn}'$ could also be induced by a change in a coordinate system that does not disturb the approximation procedure. The physical meaning of the choice ${}_2a_m = 0$ is that we assume the existence of a coordinate system in which each of the two bodies reveals its spherical symmetry; a coordinate system in which for ${}_2m \to 0$, $\xi \to 0$, the field goes over into that defined by the Schwarzschild solution in an isotropic coordinate system. The choice of such a coordinate system is implicitly assumed by our approximation procedure. However this choice of ${}_2a_m = 0$ refers only to the beginning of our approximation procedure; therefore it does not refer to ${}_3h_{0m}$. If we replace ${}_3h_{0m}$ by

$${}_3h_{0m}' = {}_3h_{0m} + {}_3a_{0,m} \tag{9.3}$$

${}_3a_0$ being an arbitrary function of x^α, then Eq. (7.5) is fulfilled just as well. This change in ${}_3h_{0m}$ can be induced by a change in a coordinate system which does not disturb our approximation procedure. Such a change also induces a simple change in ${}_4h_{00}$ (Appendix B):

$${}_4h_{00}' = {}_4h_{00} + 2\,{}_4a_{0,0}. \tag{9.4}$$

Therefore, it would seem that the Lagrangian and with it the equations of motion would change. The expressions that change in the Lagrangian (8.4) are

$$\bar{h}_{0s}\xi^s + \tfrac{1}{2}\,\overline{{}_4h_{00}{}^*} \tag{9.5}$$

and the change induced by them in the Lagrangian is, because of (9.3) and (9.4),

$$\Delta\overline{\mathfrak{L}^*} = \overline{a_{0,s}}\,\xi^s + \overline{a_{0,0}} = \frac{d\overline{a_0}}{dt}. \tag{9.6}$$

Therefore,

$$\delta\int_{t_1}^{t_2} \Delta\overline{\mathfrak{L}^*}\,dt = \delta(\overline{a_0})\Big|_{t_1}^{t_2} = 0. \tag{9.7}$$

This means: the equations of motion are uniquely determined up to the fourth approximation by the field equations and by our approximation procedure. Neither the harmonic coordinate condition nor any other coordinate condition played any role in our derivation of the equations of motion.

10. THE GENERAL THEORY

Now we formulate the general theory,‖ according to which we proceeded in our special case and in which we found the equations of motion of the second and fourth order. Such a general theory is of little practical value, since there would be hardly any physical meaning in developing the calculations one step further. Moreover, it seems—and we discuss this later—that by proper choice of the coordinate system we can annihilate all contributions to the equations of motion beyond the fourth order. From the formal point of view it is important to know that the procedure can be pushed as far as we wish. Of course we do not know anything about its convergence.

Before we formulate the general theory let us recall what has been done here. We had the Newtonian equations of motion:

$$\frac{dt}{ds}\int {}_3\mathrm{T}^{0\nu}{}_{;\nu}\,d_{(3)}x = {}_3\overline{A^0} = 0;$$
$$\frac{dt}{ds}\int {}_4\mathrm{T}^{n\nu}{}_{;\nu}\,d_{(3)}x = {}_4\bar{A}^n = 0. \tag{10.1}$$

Since ${}_2{}^1m$ appears as a factor, we called these equations (after dividing them by ${}_2{}^1m$) the equations of the *second* order. But here, since it is multiplied by ${}_2{}^1m$, it appears as an equation of the fourth order. For this section, therefore, let us rename the order of the equations of motion calling the *Newtonian* equations of motion those of the *fourth* order and the post-Newtonian equations those of the *sixth* order. Let us also put generally

$$\frac{dt}{ds}\int {}_{2n-1}T^{0\nu}{}_{;\nu}\,d_{(3)}x = {}_{2n-1}\overline{A^0};$$
$$\frac{dt}{ds}\int {}_{2n}\mathrm{T}^{n\nu}{}_{;\nu}\,d_{(3)}x = {}_{2n}\overline{A^n}. \tag{10.2}$$

Thus for our post-Newtonian equations of motion we have

$${}_3\overline{A^0} + {}_5\overline{A^0} = 0;\quad {}_4\overline{A^n} + {}_6\overline{A^n} = 0. \tag{10.3}$$

These equations gave us

$${}^am = {}_2{}^am + {}_4{}^am;\quad {}^a\xi^s = {}_0{}^a\xi^s + {}_2{}^a\xi^s, \tag{10.4}$$

where ${}_0{}^a\xi^s$ is the motion in the Newtonian approximation. To find these equations explicitly we used the Newtonian equation of motion in ${}_6\overline{A^m}$, since the use of ${}^a\xi^s$ instead of ${}_0{}^a\xi^s$ would give a contribution of the 8th order to the equations of motion. We express this idea in symbols and write instead of (10.3):

$${}_3A^0({}_0\xi^k + {}_2\xi^k) + {}_5A^0({}_0\xi^k) = 0;$$
$${}_4A^m({}_0\xi^k + {}_2\xi^k) + {}_6A^m({}_0\xi^k) = 0. \tag{10.5}$$

Thus ${}_4A^m({}_0\xi + {}_2\xi)$ also gives a contribution of the order

‖ The ideas presented here are a few years old. A. E. Scheidegger [Revs. Modern Phys. 25, 451 (1953)] refers to them in Sec. 5 stating that they were suggested by me. I found a more explicit formulation of similar ideas in a thesis by B. Rameswararao (thesis, Banares Hindu University, 1955). An alternative general theory was given in a paper by Plebański and myself [Bull. Acad. Polon. III, 4, 755 (1956)].

six. The field was solved so as to obtain (10.4) in

m, n up to the order ${}_2h_{mn}({}_0\xi)$,
$0, m$ up to the order ${}_3h_{0m}({}_0\xi)$,
$0, 0$ up to the order ${}_4h_{00}({}_0\xi)$.

Suppose that we wish to go one step further. We then have the equations of motion,

$$\begin{aligned} &{}_3A^0({}_0\xi+{}_2\xi+{}_4\xi)+{}_5A^0({}_0\xi+{}_2\xi)+{}_7A^0({}_0\xi)=0,\\ &{}_4A^m({}_0\xi+{}_2\xi+{}_4\xi)+{}_6A^m({}_0\xi+{}_2\xi)+{}_8A^m(\xi)=0. \end{aligned} \tag{10.6}$$

In ${}_3A$, ${}_5A$, ${}_4A$, ${}_6A$, the argument in (10.6) is different from that in (10.5); therefore, they give contributions *up* to the eighth order. But to find ${}_8A^m({}_0\xi)$ we have to know

$${}_4h_{mn},\ {}_5h_{0m},\ {}_6h_{00}, \tag{10.7}$$

all functions of ${}_0\xi$. Thus simply denoting

$$\mathbf{Q}^{\alpha\beta}=-8\pi(\mathbf{T}^{\alpha\beta}-\tfrac{1}{2}g^{\alpha\beta}\mathbf{T}) \tag{10.8}$$

we have to solve the equations (omitting the a's above the ξ's):

$$\begin{aligned} &{}_2\mathbf{R}^{mn}({}_0\xi+{}_2\xi)+{}_4\mathbf{R}^{mn}({}_0\xi)={}_2\mathbf{Q}^{mn}({}_0\xi+{}_2\xi)+{}_4\mathbf{Q}^{mn}({}_0\xi),\\ &{}_3\mathbf{R}^{0m}({}_0\xi+{}_2\xi)+{}_5\mathbf{R}^{0m}({}_{\,}\xi)={}_3\mathbf{Q}^{0m}({}_0\xi+{}_2\xi)+{}_5\mathbf{Q}^{0m}({}_0\xi),\\ &{}_2\mathbf{R}^{00}({}_0\xi+{}_2\xi+{}_4\xi)+{}_4\mathbf{R}^{00}(\xi_0+{}_2\xi)+{}_6\mathbf{R}^{00}({}_0\xi)\\ &\quad={}_2\mathbf{Q}^{00}({}_0\xi+{}_2\xi+{}_4\xi)+{}_4\mathbf{Q}^{00}({}_0\xi+{}_2\xi)+{}_6\mathbf{Q}^{00}({}_0\xi). \end{aligned} \tag{10.9}$$

This would seem to be an impossible task, since to solve the last Eq. (10.9) we would have to know ${}_4\xi$, which we wish to find by Eq. (10.6). However, this is not so, because we have

$${}_2\mathbf{R}^{00}=-\tfrac{1}{2}\,{}_2h_{00,\,ss}=-4\pi\sum_{a=1}^{p}{}^am\,{}^a\delta={}_2Q^{00} \tag{10.10}$$

for *arbitrary* motion. Thus we may rewrite the last equation (10.9):

$${}_4\mathbf{R}^{00}({}_0\xi+{}_2\xi)+{}_6\mathbf{R}^{00}({}_0\xi)={}_4\mathbf{Q}^{00}({}_0\xi+{}_2\xi)+{}_6\mathbf{Q}^{00}({}_0\xi). \tag{10.11}$$

Collecting here the expressions of the sixth order we find ${}_6h_{00}({}_0\xi)$. Similarly we can find ${}_5h_{0m}$ and ${}_4h_{mn}$. Thus we can push the approximation one step further.

We can now formulate the general theory. In the development of $g_{\alpha\beta}$, R, T, we took into account *arbitrary* motion. Under this assumption we developed, say

$$T^{mn}={}_4T^{mn}+{}_6T^{mn}+{}_8T^{mn}+\cdots. \tag{10.12}$$

But, instead of *arbitrary* motion, let us put into the arguments *certain* motion developed into power series:

$$\xi={}_0\xi+{}_2\xi+{}_4\xi+\cdots \tag{10.13}$$

and write, say,

$${}_{2s}T({}_0\xi+{}_2\xi+{}_4\xi). \tag{10.14}$$

If developed properly, the above expression gives contributions of the order $2s+2$ and $2s+4$. Assume that we have solved the equations of motion of the order $2r$

$$\begin{aligned} &{}_3A^0({}_0\xi+{}_2\xi+\cdots{}_{2r-4}\xi)+{}_5A^0({}_0\xi+\cdots{}_{2r-6}\xi)\\ &\qquad+\cdots{}_{2r-1}A^0({}_0\xi)=0,\\ &{}_4A^m({}_0\xi+\cdots{}_{2r-4}\xi)+{}_6A^m({}_0\xi+\cdots{}_{2r-6}\xi)\\ &\qquad+\cdots{}_{2r}A^m({}_0\xi)=0. \end{aligned} \tag{10.15}$$

This means, that we have solved the field equations in

m, n up to the order ${}_{2r-4}h_{mn}$,
$0, m$ up to the order ${}_{2r-3}h_{0m}$,
$0,0$ up to the order ${}_{2r-2}h_{00}$.

Now we wish to solve the equation of motion of order $2r+2$:

$$\begin{aligned} &{}_3A^0({}_0\xi+\cdots{}_{2r-2}\xi)+\cdots{}_{2r+1}A^0({}_0\xi)=0,\\ &{}_4A^m({}_0\xi+\cdots{}_{2r-2}\xi)+\cdots{}_{2r+2}A^m({}_0\xi)=0. \end{aligned} \tag{10.16}$$

The arguments in ${}_3A^0$, $\cdots{}_{2r-1}A^0$, ${}_4A^m$, $\cdots{}_{2r}A^m$ are different in (10.16) from those in (10.15). Therefore, they give contributions of order $2r+1$ and $2r+2$. But to find ${}_{2r+2}A^m({}_0\xi)$ we have to know

$${}_{2r-2}h_{mn},\ {}_{2r-1}h_{0m},\ {}_{2r}h_{00}.$$

We have, therefore, to solve the equations

$$\begin{aligned} &{}_2\mathbf{R}^{mn}({}_0\xi+\cdots{}_{2r-4}\xi)+\cdots{}_{2r-2}\mathbf{R}^{mn}({}_0\xi)\\ &\qquad={}_2\mathbf{Q}^{mn}({}_0\xi+\cdots{}_{2r-4}\xi)+\cdots{}_{2r-2}\mathbf{Q}^{mn}({}_0\xi),\\ &{}_3\mathbf{R}^{0m}({}_0\xi+\cdots{}_{2r-4}\xi)+\cdots{}_{2r-1}\mathbf{R}^{0m}({}_0\xi)\\ &\qquad={}_3\mathbf{Q}^{0m}({}_0\xi+\cdots{}_{2r-4}\xi)+\cdots{}_{2r-1}\mathbf{Q}^{0m}({}_0\xi),\\ &{}_2\mathbf{R}^{00}({}_0\xi+\cdots{}_{2r-2}\xi)+\cdots{}_{2r}\mathbf{R}^{00}({}_0\xi)\\ &\qquad={}_2\mathbf{Q}^{00}({}_0\xi+\cdots{}_{2r-2}\xi)+\cdots{}_{2r}\mathbf{Q}^{00}({}_0\xi). \end{aligned} \tag{10.17}$$

Everywhere in these equations, with the exception of ${}_2\mathbf{R}^{00}$ and ${}_2\mathbf{Q}^{00}$, we substitute the motion already known. However, ${}_2\mathbf{R}^{00}={}_2\mathbf{Q}^{00}$ for an *arbitrary* motion. Then we can replace the last equations in (10.17) by

$$\begin{aligned} &{}_4\mathbf{R}^{00}({}_0\xi+\cdots{}_{2r-4}\xi)+\cdots{}_{2r}\mathbf{R}^{00}({}_0\xi)\\ &\qquad={}_4\mathbf{Q}^{00}({}_0\xi+\cdots{}_{2r-4}\xi)+\cdots{}_{2r}\mathbf{Q}^{00}({}_0\xi). \end{aligned}$$

In ${}_{2r-2}\mathbf{R}^{mn}({}_0\xi)$ the expression ${}_{2r-2}h_{mn}$ appears, for the first time. In ${}_{2r-1}\mathbf{R}^{0m}$ the expression ${}_{2r-1}h_{0m}$ appears and finally in ${}_{2r}\mathbf{R}^{00}$ the expression ${}_{2r}h_{00}$ appears. Collecting all the contributions of highest order in these equations and putting them equal to zero we can find ${}_{2r-2}h_{mn}$, ${}_{2r-1}h_{0m}$, ${}_{2r}h_{00}$; for ${}_{2r}h_{00}$ the equation is purely a Poisson equation!

Looking back at Eqs. (5.8), we see that if ${}_{2r-1}h_{0m}$, ${}_{2r}h_{mn}$ is a solution of (10.17), then

$$\begin{aligned} &{}_{2r-1}h_{0m}'={}_{2r-1}h_{0m}+{}_{2r-1}a_{0,\,m},\\ &{}_{2r}h_{mn}'={}_{2r}h_{mn}+{}_{2r}a_{m,\,n}+{}_{2r}a_{n,\,m} \end{aligned} \tag{10.18}$$

is also a solution.

For example, let us put $r=2$; that is,

$${}_3h_{0m}'={}_3h_{0m}+{}_3a_{0,\,m},\quad {}_4h_{mn}'={}_4h_{mn}+{}_4a_{m,\,n}+{}_4a_{n,\,m}.$$

The choice of these functions can always be achieved by a coordinate transformation from a coordinate

system in which the a's equal zero. Then in the equations of motion of the 8th order, the derivatives of these four functions will appear. Generally, they can be so chosen as to annihilate the expressions of the 8th order in the equations of motion. But it is difficult to judge whether such a coordinate system would have any physical meaning. In any case, up to the post-Newtonian approximation, the choice of the coordinate system does not play any role as long as we stick to our approximation procedure by which its beginning is determined, that is, the choice of ${}_2h_{00}$ and ${}_2h_{mn}$.

APPENDIX A¶

To distinguish between *our* δ function and *Dirac's* δ function we shall here denote the former by δ_1. Our aim is to give a "realistic" model, showing how to construct a sequence of $\delta_1(\epsilon)$ so that $\delta_1=\lim_{\epsilon\to 0}\delta_1(\epsilon)$ and such that for every ϵ:

$$\int_\infty \delta_1(\epsilon)d_{(3)}x=1;\quad \int \delta_1(\epsilon)r^{-p}d_{(3)}x=0;\qquad p=1,2\cdots k. \tag{A.1}$$

Such a model can be gained from a model $\delta(\epsilon)$ of an ordinary Dirac δ function satisfying the following conditions:

$$\delta(\epsilon)=\delta(\epsilon,r)=\epsilon^{-3}\Delta(r/\epsilon), \tag{A.2}$$

where $\Delta(r/\epsilon)$ is such that

$$\frac{1}{4\pi}D^{(p)}=\int_0^\infty \Delta(z)z^{-p+2}dz;\quad p=1,2\cdots k \tag{A.3}$$

always exists, and

$$D^{(0)}=\int\delta(\epsilon)d_{(3)}x=4\pi\int_0^\infty z^2\Delta(z)dz=1. \tag{A.4}$$

If $\delta(\epsilon)$ does not have this property, it can be made to have it by multiplying it by $(r/\epsilon)^k$ and renormalizing.** Thus with such δ's we can form the model of our δ_1 function in the following way:

$$\delta_1(\epsilon,r)=\frac{1}{k!}\left(\frac{\partial}{\partial\epsilon}\right)^k\left[\epsilon^{k-3}\Delta\left(\frac{r}{\epsilon}\right)\right]. \tag{A.5}$$

We have to show that such a choice of δ_1, satisfies (A.1). To do so, let us start with the first equation (A.1):

$$\int\delta_1 d_{(3)}x=\frac{4\pi}{k!}\left(\frac{\partial}{\partial\epsilon}\right)^k\epsilon^k\int_0^\infty\Delta(z)z^2dz=D^{(0)}=1. \tag{A.6}$$

Now as to the second equation (A.1):

$$\int\delta_1 r^{-p}d_{(3)}x=\frac{4\pi}{k!}\left(\frac{\partial}{\partial\epsilon}\right)^k\epsilon^{k-p}\int_0^\infty\Delta(z)z^{2-p}dz$$
$$=\frac{1}{k!}\left(\frac{\partial}{\partial\epsilon}\right)^k\epsilon^{k-p}D^{(p)}=0, \tag{A.7}$$

for every integer p, if

$$1\le p\le k.$$

Thus the δ_1's defined by (A.5) satisfy (A.1).

This procedure can easily be generalized. We introduce the modified Dirac functions δ_2 from the conditions

$$\int\delta_2 r^{-p}d_{(3)}x=\omega_{(p)};\quad p=1,2\cdots k, \tag{A.8}$$

where the ω are arbitrarily *prescribed* numbers. The realistic $\delta_2(\epsilon)$ satisfying (A.8) in the limit $\epsilon\to 0$ is the following:

$$\delta_2(\epsilon)=\sum_{s=0}^{k}(D^{(k-s)})^{-1}\frac{\omega_{(k-s)}}{s!}\left(\frac{\partial}{\partial\epsilon}\right)^s\left[\epsilon^{k-3}\Delta\left(\frac{r}{\epsilon}\right)\right]. \tag{A.9}$$

We find

$$\int\delta_2(\epsilon)r^{-p}d_{(3)}x$$
$$=4\pi\sum_{s=0}^{k}(D^{(k-s)})^{-1}\frac{\omega_{(k-s)}}{s!}\left(\frac{\partial}{\partial\epsilon}\right)^s\epsilon^{k-p}\int_0^\infty z^{-p+2}\Delta(z)dz$$
$$=\sum_{s=0}^{k}(D^{(k-s)})^{-1}\frac{\omega_{(k-s)}}{s!}\left(\frac{\partial}{\partial\epsilon}\right)^s\epsilon^{k-p}D^{(p)}. \tag{A.10}$$

This is different from zero and finite for $\epsilon\to 0$, only for $s=k-p$. We have:

$$\int\delta_2 r^{-p}d_{(3)}x=\omega_{(p)}+O(\epsilon)$$

and for $\epsilon\to 0$ we have (A.8).

The use of Dirac's functions requires the prescription for the values of $\omega_{(p)}$'s. The one used in this paper is the most convenient for our purpose; it requires $\omega_{(p)}=0$; $p=1,2\cdots k$.

APPENDIX B

R_{00} *up* to the fourth approximation equals

$$R_{00}=-\tfrac{1}{2}\varphi_{,ss}-\tfrac{1}{2}\,{}_4h_{00,ss}+{}_4h_{0s,0s}-\tfrac{3}{2}\varphi_{,00}+\tfrac{1}{2}\varphi_{,s}\varphi_{,s}-\tfrac{1}{2}\varphi\varphi_{,ss}. \tag{B.1}$$

Therefore

$$R^{00}=R_{00}(1-2h_{00})=R_{00}(1-2\varphi),$$
$$\mathbf{R}^{00}=R^{00}\sqrt{-g}=(1-\varphi)R^{00}=R_{00}(1-3\varphi). \tag{B.2}$$

In (B.2) exceptionally $g=|g_{\alpha\beta}|$. We have

$$\mathbf{R}^{00}=-\tfrac{1}{2}\varphi_{,ss}+\varphi\varphi_{,ss}+\tfrac{1}{2}\varphi_{,s}\varphi_{,s}+\tfrac{1}{2}\varphi_{,00}-\tfrac{1}{2}\,{}_4h_{00,ss}. \tag{B.3}$$

¶ This is an abbreviated and changed version of two papers written in collaboration with J. Plebański.[9]

** E.g., Let us take $\delta(\epsilon)=(2\pi)^{-\frac{3}{2}}\epsilon^{-3}\exp(-\tfrac{1}{2}r^2\epsilon^{-2})$, that is $\Delta(z)=(2\pi)^{-\frac{3}{2}}\exp(-\tfrac{1}{2}z^2)$. Such $\Delta(z)$ shall be changed into

$$\Delta(z)=z^k(2\pi)^{-1}2^{-(k+3/2)}\left[\Gamma\left(\frac{k+3}{2}\right)\right]^{-1}\exp(-\tfrac{1}{2}z^2).$$

The right-hand side of our gravitational equations

$$\mathbf{R}^{\alpha\beta} = -8\pi(\mathbf{T}^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}\mathbf{T}) \quad \text{(B.4)}$$

is up to the fourth order for the "zero-zero" component:

$$-8\pi({}_2\mathbf{T}^{00} - \tfrac{1}{2}(\eta^{00} - \varphi)\,{}_2\mathbf{T}^{00}(\eta^{00} + \varphi) + {}_4\mathbf{T}^{00} - \tfrac{1}{2}\,{}_4\mathbf{T}^{00} + \tfrac{1}{2}\,{}_4\mathbf{T}^{ss}) = -4\pi({}_2\mathbf{T}^{00} + {}_4\mathbf{T}^{00} + {}_4\mathbf{T}^{ss}). \quad \text{(B.5)}$$

Because

$${}_4{}^1m = \tfrac{1}{2}\,{}_2{}^1m\dot{\xi}^s\dot{\xi}^s + \frac{{}^1m\,{}^2m}{r}, \quad {}_4{}^2m = \tfrac{1}{2}\,{}_2{}^2m\dot{\eta}^s\dot{\eta}^s + \frac{{}^1m\,{}^2m}{r},$$

we have for the right-hand side:

$$-4\pi\left({}_2{}^1m\,{}^1\delta + {}_2{}^2m\,{}^2\delta + \tfrac{3}{2}\,{}^1m\dot{\xi}^s\dot{\xi}^{s\,1}\delta + \tfrac{3}{2}\,{}^2m\dot{\eta}^s\dot{\eta}^s\,{}^2\delta + \frac{{}^1m\,{}^2m}{r}{}^1\delta + \frac{{}^2m\,{}^1m}{r}{}^2\delta\right). \quad \text{(B.6)}$$

On both sides we take only expressions (*a*) of the order four, (*b*) those that give a finite contribution to $\overline{{}_4h_{00,m}}$. Thus this relevant part of ${}_4h_{00}$ satisfies the equation

$${}_4h_{00,ss} = 2fg_{,ss} + 2gg_{,ss} + g_{,00} + g_{,s}g_{,s} + 8\pi\left(\tfrac{3}{2}\,{}_2{}^2m\dot{\eta}^s\dot{\eta}^s\,{}^2\delta + \frac{{}^2m\,{}^1m}{r}{}^2\delta\right). \quad \text{(B.7)}$$

Because

$$g_{,ss} = 8\pi\,{}^2m\,{}^2\delta; \quad f = -2\,{}^1m({}^1r)^{-1}, \quad \text{(B.8)}$$

we have for the contribution of ${}_4C$ to ${}_4h$ coming from the first two expressions:

$${}_4C_{,ss} = -4\pi\,{}^2\delta a. \quad \text{(B.9)}$$

Generally, the solution of

$${}_4C_{,ss} = -4\pi\,{}^2\delta a \quad \text{(B.10)}$$

is

$${}_4C = {}^2\bar{a}({}^2r)^{-1}, \quad \text{(B.11)}$$

where

$${}^2\bar{a} = \int_{{}^2\Omega(3)} a\,{}^2\delta d_{(3)}x. \quad \text{(B.12)}$$

Thus in our case:

$${}_4C = \frac{4\,{}^1m\,{}^2m}{r\,{}^2r}. \quad \text{(B.13)}$$

Therefore,

$${}_4h_{00,ss} = g_{,00} + g_{,s}g_{,s} + 8\pi\left(\tfrac{3}{2}\,{}^2m\dot{\eta}^s\dot{\eta}^s\,{}^2\delta - \frac{{}^1m\,{}^2m}{r}{}^2\delta\right). \quad \text{(B.14)}$$

Finally, therefore, we have

$${}_4h_{00} \to -{}^2m\,{}^2r_{,00} + 2({}^2m)^2({}^2r)^{-2} - 3\,{}^2m\dot{\eta}^s\dot{\eta}^s({}^2r)^{-1} + 2\,{}^1m\,{}^2m({}^2rr)^{-1}. \quad \text{(B.15)}$$

Let us now generalize ${}_4h_{00}$ for *three* particles, again looking only for expressions which give a contribution to ${}_4h_{00,m}$. The only nontrivial expressions of this kind are those proportional to ${}^2m\,{}^3m$.

We denote by ${}^{(ab)}r$, the "distance" between the a and b particle††:

$$({}^{(ab)}r)^2 = ({}^a\xi^s - {}^b\xi^s)({}^a\xi^s - {}^b\xi^s), \quad \text{(B.16)}$$

and ask: what contributions to (B.4) come from the third particle and are proportional to ${}^2m\,{}^3m$? We now have

$$\varphi = {}_2h_{00} = f + g + k,$$
$$f = -2\,{}^1m({}^1r)^{-1}; \quad g = -2\,{}^2m({}^2r)^{-1}; \quad k = -2\,{}^3m({}^3r)^{-1}. \quad \text{(B.17)}$$

Then the additional expressions for which we look in (B.3) are:

$$\tfrac{1}{2}(gk)_{,ss} + \tfrac{1}{2}gk_{,ss} + \tfrac{1}{2}kg_{,ss} - \tfrac{1}{2}s_{00,ss} \quad \text{(B.18)}$$

where s_{00} denotes the additional expression in ${}_4h_{00}$. The additional expressions in (B.5) because of (B.6) are

$$-4\,{}^2m\,{}^3m({}^{(23)}r)^{-1}({}^2\delta + {}^3\delta). \quad \text{(B.19)}$$

Therefore, the additional expression to ${}_4h_{00,ss}$ is:

$$s_{00,ss} \sim 8\pi\,{}^2m\,{}^3m({}^{(23)}r)^{-1}({}^2\delta + {}^3\delta) + (gk)_{,ss} + gk_{,ss} + kg_{,ss}. \quad \text{(B.20)}$$

From this we find

$$s_{00} = 2\,{}^2m\,{}^3m[({}^{(23)}r)^{-1}({}^3r)^{-1} + ({}^{(32)}r)^{-1}({}^2r)^{-1} + 2({}^2r\,{}^3r)^{-1}]. \quad \text{(B.21)}$$

Therefore, as we see, s_{00} does not depend explicitly on ${}^1\xi$. Therefore,

$$\overline{s_{00,m}} = \overline{s_{00}}_{,m} \quad \text{(B.22)}$$

and

$$\tfrac{1}{2}\,\overline{s_{00}} - \tfrac{1}{4}\,\overline{gk} = {}^2m\,{}^3m[({}^{(23)}r\,{}^{(31)}r)^{-1} + ({}^{(23)}r\,{}^{(21)}r)^{-1} + ({}^{(12)}r\,{}^{(13)}r)^{-1}]. \quad \text{(B.23)}$$

The last very simple question with which we shall deal here is the change from ${}_3h_{0m}$ to ${}_3h_{0m}{}'$:

$${}_3h_{0m}{}' = {}_3h_{0m} + {}_3a_{0,m}. \quad \text{(B.24)}$$

Putting this in (B1), we have

$${}_4h_{00}{}' = {}_4h_{00} + 2\,{}_4a_{0,0} \quad \text{(B.25)}$$

which is identical with (9.4).

†† Previously ${}^{(12)}r = r$.

Correspondence in the Generalized Theory of Gravitation*

BEHRAM KURŞUNOĞLU
Turkish General Staff, Scientific Advisory Board, Ankara, Turkey

I. INTRODUCTION

THE success of general relativity as a physical theory has made it necessary to investigate further possibilities of taking over more of the geometrical concepts for the description of natural phenomena. In its present form (Einstein's 1915 theory) general relativity is an expression of field concepts in an incomplete sense. It describes (i) gravitational phenomena in the large, and (ii) it gives the equations of motion of singularities in an external field; it does not say anything about (i) gravitational phenomena in the small, and (ii) electric charges and their interaction with fields.

The most important consequence of the nonlinearity of the gravitational field equations is the fact that they contain the law of motion. In this sense general relativity together with Maxwell's field constitutes a well-defined but an incomplete theory. In this paper the incompleteness of the theory is not taken in the sense of its as yet unquantized state. We believe the nonlinearity of the theory is more important than its possible quantized forms. It is somewhat difficult to understand the principle of equivalence in the quantized version of general relativity. The quantum nature of all atomic phenomena may be expected to result from an unquantized nonlinear theory obtained by a consistent and well-defined generalization of general relativity.

In order to test the internal consistency of a generalized theory of gravitation one can advance a correspondence argument like in quantum theory. One anticipates the reduction of the total field to general relativistic field equations on which the generalized theory is to be based. The trivial alternative in the correspondence argument is to abolish the electromagnetic field in the equations of the generalized theory that results in the free field equations of general relativity without Maxwell's field. The second alternative is obtained by means of a fundamental length in the generalized theory, the vanishing of which results in the field equations of general relativity in the presence of electromagnetic field. Of the two alternatives the latter is certainly more satisfactory in that it does not end up with a pure gravitational field which it wanted to unify with electricity from the beginning. In terms of universal constants the special relativity is based on the constancy of the velocity of light c and quantum theory on the finite value of Planck's constant h. By making c tend to infinity one obtains the nonrelativistic laws and making h tend to zero one gets classical description. In addition to the universal constants c and γ (gravitational constant) in the generalized theory a third constant is needed to formulate a correspondence argument for the theory. The latter role is played by a fundamental constant r_0 in the author's[1] version of the generalized theory of gravitation, which in the following will be referred to as (I).

In the limit of $r_0=0$ we obtain the equations of general relativity with the electromagnetic field as the source of gravitation plus Maxwell's equations for charge free fields. The existence of charges is linked up with a nonvanishing fundamental length just as the existence of spin arises from a finite h.

II. DISCUSSION OF THE FIELD EQUATIONS

The total field is described by a nonsymmetric tensor $g=a_{\alpha\beta}+\varphi_{\alpha\beta}$. The physical interpretation of symmetric and antisymmetric parts will result from a correspondence argument. The metric of the total field contains both symmetric and antisymmetric variables and it can be measured by using standard measuring rods and clocks. In (I) the metric was defined by

$$b^{\alpha\beta}=\mathfrak{g}^{\underline{\alpha\beta}}/(-\mathrm{Det}\,\mathfrak{g}^{\underline{\alpha\beta}})^{-\frac{1}{2}}, \qquad \text{(II.1)}$$

where

$$\mathfrak{g}^{\underline{\alpha\beta}}=\sqrt{(-g)}\,a^{\alpha\beta}\left[\frac{\delta^{\alpha}_{\mu}(1+\Omega)-\varphi^{\alpha\nu}\varphi_{\mu\nu}}{1+\Omega-\Lambda^2}\right]$$

$$=\sqrt{-g}\,a^{\alpha\beta}[\lambda\delta^{\alpha}_{\mu}-\rho\varphi^{\alpha\nu}\varphi_{\mu\nu}]. \qquad \text{(II.2)}$$

Hence,

$$b=\frac{g^2}{a}\,\mathrm{Det}(\lambda\delta^{\alpha}_{\mu}-\rho\varphi^{\alpha\nu}\varphi_{\mu\nu})$$

$$=\frac{g^2}{a}(\lambda^4-\rho\lambda^3P_1+\lambda^2\rho^2P_2-\lambda\rho^3P_3+\rho^4P_4), \qquad \text{(II.3)}$$

where

$$P_1=c^{\alpha}{}_{\alpha}=2\Omega,$$

$$P_2=\frac{1}{2!}\delta^{\alpha\beta}_{\mu\nu}c^{\mu}{}_{\alpha}c^{\nu}{}_{\beta}=\Omega^2-2\Lambda^2,$$

$$P_3=\frac{1}{3!}\delta^{\alpha\beta\gamma}_{\mu\nu\sigma}c^{\mu}{}_{\alpha}c^{\nu}{}_{\beta}c^{\sigma}{}_{\gamma}=-2\Omega\Lambda^2,$$

$$P_4=\frac{1}{4!}\delta^{\alpha\beta\gamma\delta}_{\mu\nu\rho\sigma}c^{\mu}{}_{\alpha}c^{\nu}{}_{\beta}c^{\rho}{}_{\gamma}c^{\sigma}{}_{\delta}=\mathrm{Det}(\varphi^{\alpha\nu}\varphi_{\beta\nu})=\Lambda^4,$$

---

* Reported at the Conference on Gravitational Theory, Chapel Hill, North Carolina, January 18–23, 1957.

[1] B. Kurşunoğlu, Phys. Rev. 88, 1369 (1952).

and

$$\overset{\alpha}{c}_{\beta}=\varphi^{\alpha\mu}\varphi_{\beta\mu}.$$

In deriving these relations we use the identities

$$a^{\mu\nu}\varphi_{\alpha\nu}\varphi_{\beta\mu}\varphi^{\beta\gamma}=\overset{\mu}{\varphi_{\alpha\cdot}}\varphi_{\beta\mu}\varphi^{\beta\gamma}=\Omega\overset{\gamma}{\varphi_{\alpha\cdot}}+\Lambda \overset{\gamma}{f_{\alpha\cdot}},$$

$$\varphi^{\mu\rho}\varphi_{\nu\rho}\varphi^{\nu\sigma}\varphi_{\mu\sigma}=2\Omega^2+4\Lambda^2,$$

$$\varphi^{\mu\rho}\varphi_{\nu\rho}\varphi^{\nu\sigma}\varphi_{\lambda\sigma}\varphi^{\sigma\gamma}\varphi_{\mu\gamma}=2\Omega^3+6\Omega\Lambda^2,$$

$$\varphi^{\mu\alpha}\varphi_{\sigma\alpha}\varphi^{\sigma\beta}\varphi_{\rho\beta}\varphi^{\rho\gamma}\varphi_{\nu\gamma}\varphi^{\nu\delta}\varphi_{\mu\delta}=2\Omega^4+4\Lambda^4+8\Omega^2\Lambda^2,$$

$$\varphi^{\alpha\mu}\varphi_{\beta\mu}=\overset{\alpha}{\delta_{\beta}}\Omega+f^{\alpha\mu}f_{\beta\mu},$$

$$f^{\alpha\mu}\varphi_{\beta\mu}=\overset{\alpha}{\delta_{\beta}}\Lambda$$

and

$$\varphi_{\alpha\beta}=\tfrac{1}{2}\sqrt{(-a)}\,\epsilon_{\alpha\beta\mu\nu}f^{\mu\nu},$$

$$\varphi^{\alpha\beta}=-\frac{1}{2}\frac{1}{\sqrt{(-a)}}\epsilon^{\alpha\beta\mu\nu}f_{\mu\nu}.$$

Hence (II.3) leads to

$$\frac{ab}{g^2}=\frac{1}{(1+\Omega-\Lambda^2)^2}$$

or

$$b=a. \tag{II.4}$$

The contravariant metric thus given by

$$b^{\alpha\beta}=\frac{a^{\alpha\beta}(1+\Omega)-\varphi^{\alpha\mu}\varphi^{\beta}{}_{\cdot\mu}}{(1+\Omega-\Lambda^2)^{\frac{1}{2}}}. \tag{II.5}$$

The covariant metric tensor can be obtained from

$$b_{\alpha\mu}=\frac{a}{3!}\epsilon_{\alpha\beta\gamma\delta}\epsilon_{\mu\nu\rho\sigma}b^{\beta\nu}b^{\gamma\rho}b^{\delta\sigma}$$

as

$$b_{\alpha\beta}=\frac{a_{\alpha\beta}+\overset{\mu}{\varphi_{\alpha\cdot}}\varphi_{\beta\mu}}{(1+\Omega-\Lambda^2)^{\frac{1}{2}}},\quad b^{\alpha\mu}b_{\beta\mu}=\overset{\alpha}{\delta_{\beta}}. \tag{II.6}$$

The determinant of $\overset{\alpha\beta}{\mathfrak{g}}$ is

$$\mathrm{Det}(\overset{\alpha\beta}{\mathfrak{g}})=-a\Lambda^2=\mathrm{Det}(f_{\mu\nu}).$$

We may interpret the geometrical entity $g_{\alpha\beta}$ as a physical tensor by introducing a critical field q (see I) given by

$$q=pc^2/(2\gamma)^{\frac{1}{2}}=c^2/r_0(2\gamma)^{\frac{1}{2}} \tag{II.7}$$

and write

$$g_{\alpha\beta}=a_{\alpha\beta}+r_0(2\gamma)^{\frac{1}{2}}\varphi_{\alpha\beta}/c^2. \tag{II.8}$$

The theory is invariant with respect to the sign of r_0. Both $g_{\alpha\beta}$ and $g_{\beta\alpha}=a_{\alpha\beta}-r_0(2\gamma)^{\frac{1}{2}}/c^2\varphi_{\alpha\beta}$ satisfy the same set of equations. Another important invariance of the theory refers to the λ invariance defined by

$$\overset{*}{\Gamma}{}^{\gamma}_{\alpha\beta}=\Gamma^{\gamma}_{\alpha\beta}+\delta^{\gamma}_{\alpha}\lambda_{,\beta}, \tag{II.9}$$

where λ is a scalar. The most important significance of the λ invariance lies in its exclusion of complex field variables $g_{\alpha\beta}$ and complex $\Gamma^{\gamma}_{\alpha\beta}$. As we shall see (in the next section) the complex $g_{\alpha\beta}$ would lead to a velocity of light exceeding the constant c.

The field equations are

$$R_{\underline{\alpha\beta}}=-p^2\left[a_{\alpha\beta}-\frac{a_{\alpha\beta}+(2\gamma/p^2c^4)\overset{\mu}{\varphi_{\alpha\cdot}}\varphi_{\beta\mu}}{[(1+(2\gamma/p^2c^4)\Omega-(4\gamma^2/p^4c^8)\Lambda^2]^{\frac{1}{2}}}\right], \tag{II.10}$$

$$R_{\underset{\smile}{\alpha\beta},\gamma}+R_{\underset{\smile}{\beta\gamma},\alpha}+R_{\underset{\smile}{\gamma\alpha},\beta}=-p^2I_{\alpha\beta\gamma}, \tag{II.11}$$

$$\psi_{\alpha\beta,\gamma}+\psi_{\beta\gamma,\alpha}+\psi_{\gamma\alpha,\beta}=0, \tag{II.12}$$

where the negative sign before p^2 is taken in anticipation of a comparison of the theory with general relativity and where

$$\psi_{\alpha\beta}=-\tfrac{1}{2}\epsilon_{\alpha\beta\mu\nu}\overset{\mu\nu}{\underset{\smile}{\mathfrak{g}}}=\frac{f_{\alpha\beta}+\dfrac{2\gamma}{p^2c^4}\varphi_{\alpha\beta}\Lambda}{\left(1+\dfrac{2\gamma}{p^2c^4}\Omega-\dfrac{4\gamma^2}{p^4c^8}\Lambda^2\right)^{\frac{1}{2}}}. \tag{II.13}$$

The electric current is defined by

$$\mathfrak{J}^{\alpha}=\frac{1}{3!}\epsilon^{\mu\nu\rho\alpha}I_{\mu\nu\rho}$$

$$I_{\mu\nu\rho}=\varphi_{\mu\nu,\rho}+\varphi_{\nu\rho,\mu}+\varphi_{\rho\mu,\nu}. \tag{II.14}$$

If we use the algebraic equations

$$g_{\alpha\beta,\gamma}=g_{\mu\beta}\Gamma^{\mu}_{\alpha\gamma}+g_{\alpha\mu}\Gamma^{\mu}_{\gamma\beta}$$

defining the displacement field $\Gamma^{\gamma}_{\alpha\beta}$, and make p tend to infinity, the field equations reduces to

$$G_{\alpha\beta}=(2\gamma/c^4)T_{\alpha\beta}, \tag{II.15}$$

$$f_{\alpha\beta,\gamma}+f_{\beta\gamma,\alpha}+f_{\gamma\alpha,\beta}=0, \tag{II.16}$$

where

$$T_{\alpha\beta}=\overset{\mu}{f_{\alpha\cdot}}f_{\beta\mu}-\tfrac{1}{4}a_{\alpha\beta}f^{\mu\nu}f_{\mu\nu}. \tag{II.17}$$

The equations (II.11) lead to

$$\varphi_{\alpha\beta,\gamma}+\varphi_{\beta\gamma,\alpha}+\varphi_{\gamma\alpha,\beta}=0 \tag{II.18}$$

which expresses the vanishing of free electric charges in the limit of $r_0=0$. Thus the existence of free charges is linked up with a finite fundamental length r_0. We can apply the correspondence argument to the conservation laws of the total field

$$\mathfrak{T}^{\rho}_{\alpha,\rho}=0, \tag{II.19}$$

where

$$-4\pi p^2 q^{-2}\mathfrak{T}^{\alpha}_{\beta}=\tfrac{1}{2}(\mathfrak{g}^{\alpha\mu}R_{\beta\mu}+\mathfrak{g}^{\mu\alpha}R_{\mu\beta}-\delta^{\alpha}_{\beta}\mathfrak{g}^{\mu\nu}R_{\mu\nu})$$

$$+\tfrac{1}{2}(\mathfrak{g}^{\mu\nu}_{,\beta}\,\mathfrak{B}^{\alpha}_{\mu\nu}-\delta^{\alpha}_{\beta}\mathfrak{B})$$

$$=(\mathfrak{g}^{\overset{\alpha\mu}{-}}R_{\underline{\beta\mu}}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{g}^{\overset{\mu\nu}{-}}R_{\underline{\mu\nu}})$$

$$+q^{-2}(\mathfrak{g}^{\overset{\alpha\mu}{\smile}}R_{\underset{\smile}{\beta\mu}}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{g}^{\overset{\mu\nu}{\smile}}R_{\underset{\smile}{\mu\nu}})$$

$$+\tfrac{1}{2}(\mathfrak{g}^{\mu\nu}_{,\beta}\,\mathfrak{B}^{\alpha}_{\mu\nu}-\delta^{\alpha}_{\beta}\mathfrak{B}). \quad \text{(II.20)}$$

If we make p tend to infinity

$$-\mathfrak{T}^{\alpha}_{\beta}=\frac{c^4}{8\pi\gamma}(\mathfrak{G}^{\alpha\mu}G_{\beta\mu}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{G}^{\mu\nu}G_{\mu\nu})+\frac{c^4}{16\pi\gamma}t^{\alpha}_{\beta}$$

$$=\frac{1}{4\pi}(-\alpha)^{\frac{1}{2}}T^{\alpha}_{\beta}+\frac{c^4}{16\pi\gamma}t^{\alpha}_{\beta}, \quad \text{(II.21)}$$

where

$$\mathfrak{G}^{\alpha\beta}=(-\alpha)^{\frac{1}{2}}a^{\alpha\beta}$$

and

$$t^{\alpha}_{\beta}=\mathfrak{G}^{\mu\nu}_{,\beta}\,\mathfrak{L}^{\alpha}_{\mu\nu}-\delta^{\alpha}_{\beta}\mathfrak{L}$$

which is the energy momentum tensor of general relativity in the presence of electromagnetic and gravitational fields.

There is no way of obtaining Einstein's version of the theory from the present one without loss of internal consistency and the physical interpretation that follows from the correspondence argument. To make $p=0$ is as meaningless as making $c=0$ in special relativity or $h=\infty$ in quantum theory. However, one can also introduce a critical field strength q in Einstein's theory and consider the limit of q tending to infinity in which case the resulting equations are two sets of disconnected equations for gravitational field in empty space and an unfamiliar set for the antisymmetric field. One does not get Maxwell's equations.

The critical field strength is responsible for the nonlinearity of the theory. The constant r_0 implies an extended charge distribution of elementary particles and nonexistence of point charges. That the electron has an extended charge distribution can be observed experimentally with electron scattering of wave length r_0. For $r_0=10^{-13}$ cm the corresponding electron energy is about 600 Mev. Some experiments on electron scattering do actually imply an extended charge distribution.

Invariance of the theory with respect to the sign of r_0 introduces the concept of negative length. Measurement of negative length may be compared to saying that there is a fictitious space where moving clocks go faster and moving measuring rods get longer! In actual case the measured length is still positive just as an electron moving in the direction of negative time arrow has a positive energy. Under these circumstances if we contemplate the origin of mass as being from the electromagnetic field then the field must create positive and negative charges simultaneously. In this sense the electromagnetic field creates pairs of particles and does not allow for single acts of creation of particles. This interpretation makes it possible to think of the existence of negative mass along with the positive mass. The vanishing of r_0 is one reason for general relativity leaving the sign of the mass quite arbitrary.

III. LIGHT CONE

The signature of the total field as in the theory of gravitation is defined by the symmetric part of $g_{\alpha\beta}$ so that the presence of the electromagnetic field does not spoil the invariance of the propagation of light but introduces an anisotropy in the form of the light cone. Because of the result $\mathrm{Det}(b_{\alpha\beta})=a$ the requirement that $\mathrm{Det}(b_{\alpha\beta})$ is everywhere different from zero and b be negative is satisfied. The propagation of light, i.e., the interval dx^α for two neighboring events shall satisfy

$$ds^2=b_{\alpha\beta}dx^\alpha dx^\beta=0, \quad \text{(III.1)}$$

which proposition is compatible with the correspondence argument since for vanishing fundamental length it reduces to the law of propagation of light in the presence of a purely gravitational field. The motivation behind choosing $b_{\alpha\beta}$ as the metric of the total field can also be understood from the Bianchi identities

$$(\mathfrak{g}^{\overset{\alpha\mu}{-}}R_{\underline{\beta\mu}}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{g}^{\overset{\mu\nu}{-}}R_{\underline{\mu\nu}})_{||\alpha}=\tfrac{1}{2}\mathfrak{g}^{\overset{\mu\nu}{\smile}}(R_{\underset{\smile}{\mu\nu},\beta}+R_{\underset{\smile}{\nu\beta},\mu}+R_{\underset{\smile}{\beta\mu},\nu}), \quad \text{(III.2)}$$

which is the limit of $r_0=0$ reduce to

$$(\mathfrak{A}^{\alpha\mu}G_{\beta\mu}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{A}^{\mu\nu}G_{\mu\nu})_{|\alpha}=0 \quad \text{(III.3)}$$

while in the absence of charges they become

$$(\mathfrak{g}^{\overset{\alpha\mu}{-}}R_{\underline{\beta\mu}}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{g}^{\overset{\mu\nu}{-}}R_{\underline{\mu\nu}})_{||\alpha}=0. \quad \text{(III.4)}$$

Thus in the absence of charges the tensor

$$\mathfrak{g}^{\overset{\alpha\mu}{-}}R_{\underline{\beta\mu}}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{g}^{\overset{\mu\nu}{-}}R_{\underline{\mu\nu}}$$

or if the field equations are satisfied the tensor

$$\mathfrak{g}^{\overset{\alpha\mu}{-}}a_{\underline{\beta\mu}}-\tfrac{1}{2}\delta^{\alpha}_{\beta}\mathfrak{g}^{\overset{\mu\nu}{-}}a_{\underline{\mu\nu}}$$

is conserved covariantly with respect to $b_{\alpha\beta}$. The metric $b_{\alpha\beta}$ is not independent of the fundamental length r_0 and in this sense the zero interval in (III.1) is one between two "extended events" and not between two point events. This can best be seen by choosing a local coordinate system in which the mixed tensor $\mathfrak{A}^{\alpha\mu}b_{\beta\mu}$ is represented by a diagonal matrix $\rho_\alpha\delta^\alpha_\beta$. From the trans-

formation law

$$(a^{\alpha\mu}b_{\beta\mu})^* = \overset{*\alpha}{b}_\beta = \frac{\partial \overset{*}{x}^\alpha}{\partial x^\mu}\frac{\partial x^\nu}{\partial \overset{*}{x}^\beta}\overset{\mu}{b}_\nu, \tag{III.5}$$

we would like to find a transformation which will result in

$$\overset{*\alpha}{b}_\beta = \rho_\alpha \overset{\alpha}{\delta}_\beta. \tag{III.6}$$

From (III.5) and (III.6) we get

$$\frac{\partial \overset{*}{x}^\alpha}{\partial x^\mu}\rho_\alpha = \frac{\partial \overset{*}{x}^\alpha}{\partial x^\nu}\overset{\nu}{b}_\mu$$

or

$$\frac{\partial \overset{*}{x}^\alpha}{\partial x^\mu}(\overset{\mu}{b}_\beta - \overset{\mu}{\delta}_\beta \rho_\alpha) = 0. \tag{III.7}$$

These 16 equations will determine the four eigenvalues ρ_α and the transformation $\partial\overset{*}{x}^\alpha/\partial x^\beta$. The eigenvalues are found from the condition

$$\operatorname{Det}(\overset{\alpha}{b}_\beta - \overset{\alpha}{\delta}_\beta \rho_\gamma) = 0. \tag{III.8}$$

By applying the methods of Sec. II, we obtain the equation

$$\rho^4 - I_1\rho^3 + I_2\rho^2 - I_1\rho + 1 = 0, \tag{III.9}$$

where

$$I_1 = \frac{4+2\Omega}{(1+\Omega-\Lambda^2)^{\frac{1}{2}}},$$

$$I_2 = 6 + \frac{\Omega^2+4\Lambda^2}{1+\Omega-\Lambda^2} = \tfrac{1}{4}I_1^2 + 2.$$

The equation (III.9) can be written as

$$\rho^2 + \frac{1}{\rho^2} - I_1\left(\rho + \frac{1}{\rho}\right) + I_2 = 0$$

so by putting $x = \rho + 1/\rho$ it reduces to

$$x^2 - I_1 x + \tfrac{1}{4}I_1^2 = 0$$

with the roots

$$\rho = \tfrac{1}{4}I_1 \pm \tfrac{1}{2}(I_2 - 6)^{\frac{1}{2}}.$$

The other two roots are obtained by noting that the equation (III.9) is also satisfied by $1/\rho$. Thus the eigenvalues are

$$\rho_1 = \tfrac{1}{4}I_1 + \tfrac{1}{2}\surd(I_2-6) = \rho_4,$$
$$\rho_2 = \tfrac{1}{4}I_1 - \tfrac{1}{2}\surd(I_2-6) = \rho_3. \tag{III.10}$$

In the local frame as chosen, the space-time interval is

$$ds^2 = a_{\alpha\beta}\Delta x^\alpha \Delta x^\beta, \tag{III.11}$$

where

$$\Delta x^\alpha = (\rho_\alpha)^{\frac{1}{2}} dx^\alpha, \tag{III.12}$$

so there will be an electromagnetic extension of the two neighboirng point events. Because of the algebraic invariants involved in the definition of the metric it is not possible to choose a local frame in which the metric tensor will take the form of the metric of the flat space time. This is the same as saying that the principle of equivalence would not hold in the presence of an electromagnetic field.

It is convenient to express the metric $b_{\alpha\beta}$ in terms of $\psi_{\alpha\beta}$. We use the antisymmetric tensors $\psi_{\alpha\beta}$ and $\phi_{\alpha\beta}$ to record the invariants

$$\Lambda = \tfrac{1}{4}\phi^{\alpha\beta}\psi_{\alpha\beta},$$

$$Q = \tfrac{1}{2}\psi^{\alpha\beta}\psi_{\alpha\beta} = \frac{\Omega\Lambda^2 - \Omega + 4\Lambda^2}{1+\Omega-\Lambda^2}$$
$$= -\tfrac{1}{2}\phi^{\alpha\beta}\phi_{\alpha\beta}(= -\tfrac{1}{2}b^{\alpha\mu}b^{\beta\nu}\varphi_{\mu\nu}\varphi_{\alpha\beta}),$$

where

$$\phi_{\alpha\beta} = \frac{1}{\sqrt{-a}}a_{\alpha\mu}a_{\beta\nu}\overset{\mu\nu}{\breve{g}} = \frac{\varphi_{\alpha\beta} - \Lambda f_{\alpha\beta}}{(1+\Omega-\Lambda^2)^{\frac{1}{2}}}.$$

Hence

$$(1+Q-\Lambda^2)^{\frac{1}{2}} = \frac{1+\Lambda^2}{(1+\Omega-\Lambda^2)^{\frac{1}{2}}}$$

and

$$\phi^{\alpha\mu}\phi_{\beta\mu} = -\overset{\alpha}{\delta}_\beta Q + \psi^{\alpha\mu}\psi_{\beta\mu},$$

$$\varphi_{\alpha\beta} = \frac{\phi_{\alpha\beta} + \Lambda\psi_{\alpha\beta}}{(1+Q-\Lambda^2)^{\frac{1}{2}}} \tag{III.13}$$

$$f_{\alpha\beta} = \frac{\psi_{\alpha\beta} - \Lambda\phi_{\alpha\beta}}{(1+Q-\Lambda^2)^{\frac{1}{2}}}. \tag{III.14}$$

The metric takes the form

$$b_{\alpha\beta} = \frac{a_{\alpha\beta} + \overset{\mu}{\psi}_{\alpha\cdot}\psi_{\beta\mu}}{\surd(1+Q-\Lambda^2)} \tag{III.15}$$

which shows that the metrical properties of the total field is the same with respect to $\varphi_{\alpha\beta}$ and $\psi_{\alpha\beta}$ fields. The electromagnetic field is described by four vectors

$$\mathbf{D} = (\varphi_{23}, \varphi_{31}, \varphi_{12}), \quad \mathcal{H} = (\varphi_{41}, \varphi_{42}, \varphi_{43}),$$
$$\mathbf{B} = (\psi_{23}, \psi_{31}, \psi_{12}), \quad \mathcal{E} = (\psi_{14}, \psi_{24}, \psi_{34})$$

that reduce to two vectors in the correspondence limit.†

The metric can be written as

$$b_{\alpha\beta} = a_{\alpha\mu}\overset{\mu}{c}_\beta \tag{III.16}$$

where

$$\overset{\mu}{c}_\beta = \frac{\delta_\beta{}^\mu + \psi^{\mu\rho}\psi_{\beta\rho}}{(1+Q-\Lambda^2)^{\frac{1}{2}}}$$

† In connection with the recent observation on parity nonconservation, it is interesting to note that the existence of the Λ term (where $\Lambda = \mathcal{E}\cdot\mathbf{B}$) in (III.13) and (III.14) implies different field strength in the reflected coordinate system. The parity conservation must also be violated for very strong fields such as at a charged particle itself.

and

$$\text{Det}(c_\beta{}^\mu)=1.$$

Because of the form (III.16) the signature of the metric will remain unchanged by coordinate transformations. But there exists no local frame in which the measuring rods and clocks will operate without being contracted and retarded. These facts about $b_{\alpha\beta}$ make it clear that we can take over the concept of light cone from general relativity to study the propagation of $\varphi_{\alpha\beta}$ waves. By using the form (III.15) of the metric we find that the distortion of the light cone due to the finite value of the critical field is given by

$$a_{\alpha\beta}dx^\alpha dx^\beta=-q^{-2}\psi_{\alpha\cdot}{}^{\mu}\psi_{\beta\mu}dx^\alpha dx^\beta.$$

This can be written

$$\left(1+\frac{q^{-2}}{2}Q\right)a_{\alpha\beta}dx^\alpha dx^\beta=-q^{-2}T_{\alpha\beta}dx^\alpha dx^\beta \quad \text{(III.17)}$$

or

$$\left(1+\frac{q^{-2}}{2}Q\right)=-q^{-2}T_{\alpha\beta}\frac{dx^\alpha}{du}\frac{dx^\beta}{du}, \quad \text{(III.18)}$$

where

$$du^2=a_{\alpha\beta}dx^\alpha dx^\beta. \quad \text{(III.19)}$$

The result (III.18) is in complete agreement with Schrödinger's[2] calculation obtained under certain assumptions about the ψ field from the equations (II.12) and (II.16). Schrödinger's calculation is based on a special frame where $a_{\alpha\beta}$ assumes a Galilean form. He chooses the principal directions of the three-dimensional part of $T_j{}^i$ of $T_\beta{}^\alpha$ as the coordinate axes and takes $T_i{}^4$ parallel to the x_1 direction, in which case (III.18) assumes a simpler form. In (III.18) we can regard the ψ field consisting of two parts: an infinitely weak, rapidly oscillating part that represents the light wave whose propagation is to be investigated, and a beckground field with no restriction on its magnitude. For the special case of Galilean a the equation (III.18) is

$$\frac{p^2c^4}{2\gamma}(c^2-v^2)=-(\mathbf{v}\cdot\mathbf{B})^2+c^2B^2+(\boldsymbol{\mathcal{E}}\times\mathbf{v})^2-2c\mathbf{v}\cdot(\boldsymbol{\mathcal{E}}\times\mathbf{B}). \quad \text{(III.20)}$$

In the case of a purely electric field $\mathcal{E}$ the D field is given by

$$\mathbf{D}=\kappa_0\boldsymbol{\mathcal{E}}, \quad \text{(III.21)}$$

where

$$\kappa_0=\left(1-\frac{\mathcal{E}^2}{q^2}\right)^{-\frac{1}{2}}$$

can be regarded as a dielectric constant for the background field. The velocity of light v in the direction of the wave normal in the background field, independently of its polarization and frequency, is

$$v^2=\frac{c^2}{1+(\mathcal{E}^2/q^2)\sin^2\theta}. \quad \text{(III.22)}$$

If the background field $\mathcal{E}$ is very small compared to the critical field q, we have

$$v^2=c^2(\kappa_0{}^2\sin^2\theta+\cos^2\theta), \quad \text{(III.23)}$$

where θ is the angle between the wave normal and $\boldsymbol{\mathcal{E}}$. Another simple case is obtained for a background field with parallel $\mathcal{E}$ and B fields. Both results agree with Schrödinger's calculations. For complex field variables $g_{\alpha\beta}$ the result (III.22) would have a denominator smaller than 1, so the velocity of light would exceed the constant c.

IV. CONCLUSION

With the physical interpretation proposed in this paper, the concept of generalized theory of gravitation may be taken more seriously than it has been looked on in the past. The correspondence argument used in this theory puts great emphasis on the validity of the general relativity as a physical theory. The latter is either a virtue or an undesirable basis for a theory aiming at a unified description of the natural phenomena. However, one thing is quite clear now, that general theory of relativity with all its tested consequences, is the only genuine field theory in physics. It is a field theory that is independent of any mechanical constants like mass and charge, where masses are singularities in the field. The generalized theory is expected to account for electric charge and mass. One of the necessary conditions for the generalized field to represent mass and charge must come from the regularity of the field everywhere. Particles may be represented as stationary solutions of the field equations with large fluctuations confined to regions of the dimensions of r_0. Unfortunately, we have not yet been able to get exact solutions of the field equations to verify the above conjectures. One of the most important reasons for requiring exact solutions lies in the fact that the critical field may be approached at places where particles are situated. In this case we cannot disregard the nonlinearity of the theory. Strong gravitational and electromagnetic interactions may be expected to take place in regions of high fluctuations. The concept of negative length may be linked up with negative mass (or antimatter). The shape of the universe would be like a dipole with repelling poles consisting of matter and antimatter. The poles of the dipole should be rotating around the center of the dipole to maintain a steady state universe against gravitational repulsion of the poles.

It still remains to be seen in what way Planck's constant will appear in the theory, without which it would be empty of any physical content.

[2] O. Hittmair and E. Schrödinger, Commun. Dublin Inst. Advanced Studies, Series A, No. 8 (1951).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

General Relativity and the Divergence Problem in Quantum Field Theory

S. DESER*

Institute for Theoretical Physics, University of Copenhagen, Copenhagen, Denmark

I. INTRODUCTION

POSSIBLE physical consequences of general relativity for the elementary particle problem have usually been assumed to be negligible, due to the weakness of the gravitational coupling constant. It has been felt that since the effects of the new particles and the energies at which current theory becomes invalid occur very much below the domain for which gravitation becomes significant, a correct future theory will obviate the necessity of considering relativity in this context. It is presumed that, in any case, such a correct theory will differ so radically from current ideas, that arguments based on these will lose their validity. Further, general relativity is sometimes thought to be purely macroscopic in applicability, and to lose its meaning in subatomic regions where the concept of metric (or possibly metric operator) is not very transparent.

To these classic objections, several levels of reply may be given. Thus, it may be pointed out, for example, that the effect of a theory is not always felt first through its gross direct dynamical contributions. Also, general relativity provides the most natural definition of the stress-energy tensor of matter and, in that sense, underlies any theory dealing with masses and energies of fields and particles. The gravitational field is the only one which by definition is coupled to all matter fields, and always in the same way. The principle of general covariance on which the general theory is based is in no way restricted to macroscopic considerations (elementary particles are affected by gravitational fields), and it is thus necessary to explore its effects on all fields. The fact that an interval or metric component may not be simply measurable "within," say, an elementary particle is no more an argument against the relevance of relativity there than is the complex definition of electron position in an hydrogen atom an argument against the use of coordinates. It appears feasible to extend the usual quantum mechanical measurement theory to cover gravitational couplings also.

The relevance of the general covariance principle to the elementary particle domain has long been recognized by O. Klein, from whose program in this connection derive many of the arguments of this section.[1]

Such general arguments may be supplemented with very considerable quantitative ones. The whole scheme of current Lorentz-invariant field theory is plagued with divergences occurring because there is no limit to the energies involved in it, while the relevant functions do not decrease rapidly enough with energy to make the integrals exist. No satisfactory cutoff method is known which preserves the basic requirements of physical meaningfulness, and there is now strong doubt as to whether any can be found within a Lorentz-covariant framework. At this point, general relativity may be invoked. Classical considerations indicate that for any kind of matter coupled to the metric field in the Einstein way, there are limitations on the energy densities and masses which can be concentrated or built up in a given region.[2] In all such investigations, space-time loses its physically meaningful character beyond such limiting densities (or alternately, below such limiting dimensions), singularities then appearing in the solutions. Similarly, a lower bound exists on the size of wave packets built up from linear wave equations, below which gravitational self-effects enter. This bound is independent of the specific field involved; for an energy E, and a packet of linear dimension d, $E<c^4d/2\gamma$, while $E=\hbar c/d$, so that[3] $d>l_0=(2\hbar c\kappa)^{\frac{1}{2}}$. Thus, below this fundamental length $l_0\sim 10^{-32}$ cm, the packet can no longer be taken to be linear.[2] Although this distance is very small, it does correspond to a finite energy, and its effects may well be felt earlier. That l_0 is a universal constant independent of the parameters of particular fields is due to the fact that at these energies the actual masses are negligible, as will be seen below; the adjunction of κ to $\hbar$ and c defines a length or energy automatically.

Still more significant is the following. As is well known, under very wide assumptions, theories of coupled fields lead, near the light cone, to singularities in the propagators of the clothed particles and to the existence of at least some infinite renormalization constants, independent of perturbation theory.[4] All known Lorentz-invariant couplings are of this type, in particular electrodynamics and the renormalizable meson theories.

* National Science Foundation Postdoctoral Fellow.

[1] O. Klein, Kosmos 32, 33 (1954); Helv. Phys. Acta Suppl. 4, 58 (1956); and *Niels Bohr and the Development of Physics* (Pergamon Press, London, 1955), p. 96.

[2] The Schwarzchild solution of the field equations for a perfect liquid, for example, sets limits on the mass of a sphere of given radius and size for given density. General considerations as to requirements on matter to keep nonsingular metrics may be found in A. Lichnérowitz, *Théories Relativistes de la Gravitation et de l'Éléctromagnétisme* (Masson et Cie, Paris, 1955).

[3] γ is the Newtonian, κ the Einstein gravitational constant. Henceforth natural units, $\hbar=c=1$ will be used; $\kappa^{\frac{1}{2}}$ has then dimensions of a length.

[4] G. Källén, Physica 19, 850 (1953); H. Lehmann, Nuovo cimento 11, 342 (1954).

It is just in the basic conditions for the proofs of such "pessimistic" theorems that the gravitational field and its coupling differ profoundly from the matter field interactions. First, it is necessary that there exist an energy-momentum vector, P_μ, such that the usual commutation relations hold for any operator $0(x)$,

$$\partial_\mu 0(x) = i[0, P_\mu]. \tag{1}$$

This is not the case in the general theory, and is very significant there; a P_μ can only be defined for a closed system in such a way as to be independent of the choice of inner coordinates,[5] that is, in terms of over-all properties of the system, and therefore cannot yield values (or derivatives) of quantities which are connected with inner properties or regions. The nonlinearity, and consequent lack of simple translational properties, of parts of the system by themselves are here brought out; attempts to fix precisely the position of an arbitrary component of a "clothed" electron, for example, would clash with the energy density bounds mentioned earlier. Further, matter and gravitational energies are not invariantly separable.

A second, related, departure from usual theory lies in the form of the interaction terms. Contrary to the assumption always tacitly made there, the matter Lagrangian simply has no free-particle part any longer; gravitational coupling enters in the kinetic energy, in a multiplicative fashion, and one cannot really disentangle a noninteracting particle here. The equal-time matter commutation relations also depend on the position in space-time involved (and therefore on the matter distribution itself) by way of the now dynamical metric tensor.[6] The differences due to this fundamentally different mode of coupling will occur most critically at high energies, and forbid any approximations from "free-particle" just at the point where difficulties now exist. In particular, then, matter fields will no longer be representable by linear wave equations in the high-energy domain.[7] The free motion varies with the geometry of the space involved, which is in turn conditioned by the matter and its motion. It will be seen in Sec. III that the results inferred in this investigation as to the nonsingularity of propagators are based in good part on just these characteristics of gravitational coupling. It may well be that a more adequate future theory of particles will incorporate some of these aspects of the nonlinear gravitational interaction. Certainly, if it be a field theory, it must escape the pessimistic conclusions for current theory by some such channel. If so, it may be all the more instructive to see how an improvement seems to emerge here; in any case, the gravitational cutoff may remain as a backstop to such possible convergent future theories.

II. QUANTIZATION

Once gravitational interaction is introduced in the microscopic domain, it is necessary that the corresponding field be quantized; from the fact that the stress-tensor is an operator, it follows that the curvature tensor must be one as well.[8] While it is conceivable that a classical approximation for the metric quantities may suffice to secure convergence of field theory (through the above mentioned classical consideration), it is important to verify that the full quantized theory is also finite. It is outside the scope of this paper to discuss the quantization problem and the difficulties to which it may give rise when applied to the gravitational field. The method to be employed here to effect the quantization is that of Feynman.[9] It is known to be equivalent to the canonical one for the usual (quadratic Lagrangian) fields, and seems to avoid at least some of the difficulties encountered by extensions of the canonical method to this highly nonlinear problem. Whether or not the two procedures differ somewhat in this case, the Feynman method appears to provide a consistent approach. The question of which components are dynamically independent and which fully determined by the matter quantities need not, in a certain sense, be known beforehand.[10] This prescription has the further advantage, for our purposes, of exhibiting explicitly the crucial differences, stressed earlier, between matter and gravitational interactions. The general framework is as in ordinary theory: the vacuum expectation value of the operator of interest is the functional integral (suitably weighted and normalized) of a corresponding c-number quantity over all classical configurations of the system. It may be expected, roughly, that the contributions of the "scalar" and "transverse" parts of the gravitational field, though not really separable, will be to limit the density of matter in a given region for the former, and to "smear out the light cone" for the latter. This phrase refers to the expectation that the summation over all possible Riemann spaces corresponding to the various classical possibilities will average away the uniqueness of the Minkowski light cone, and with it the singularities of the propagators on such a cone. Such an effect would

[5] W. Pauli, *Encyklopädie der Mathematischen Wissenschaften* (B. G. Teubner, Leipzig, 1921), Vol. V2, pp. 539–775.

[6] For a given nonflat geometry, the situation would essentially be the same as the flat space-time one. That is, by suitable modification, the content of Eq. (1) could still be maintained.

[7] It might be thought that by writing $g_{\mu\nu} = \delta_{\mu\nu} + 0(\kappa)$ such a free part would be obtained. However such expansions in κ are invalid at high energies. The limiting process as κ approaches zero will itself be seen to proceed in a more subtle way than such an expansion of gravitational quantities in powers of κ.

[8] While such "self-consistent" approaches as starting from $R_{\mu\nu} = -\kappa\langle T_{\mu\nu}\rangle$ may be considered, they are only approximate in nature.

[9] For a recent discussion of the functional integral procedure, see P. T. Matthews and A. Salam, Nuovo cimento **2**, 120 (1955).

[10] Some of the possible difficulties of Feynman quantization of general relativity have lately been resolved: C. Misner, Revs. Modern Phys. **29**, 497 (1957), discusses the proper choice of integration variables; B. Laurent, Nuovo cimento **10**, 1445 (1956), considers the invariance of the integrals under general coordinate transformations. Laurent also has clarified other aspects of the choice of variables and of the quantization problem (private communication).

realize a possibility also remarked on by Pauli at the Bern relativity conference in 1955.[11]

The usual field theory is known to be fully determined by knowledge of all vacuum propagators, that is, vacuum expectation values of all powers of the various field operators. From these, the various S-matrix elements are directly obtainable. In our case, the strong nonlinearity makes the relations between propagators and physical quantities more complicated (so that their physical significance is less immediate); however, they still determine the theory, and their singularities, if any, are still the relevant ones.

III. MATTER PROPAGATORS

We are interested in comparing the one-particle Green's function G', corrected for both electromagnetic (or mesic) and gravitational self-effects, with S_F', the flat space electromagnetically corrected one. The latter behaves at least as singularly on the light cone as the free propagator, $S_F(p)=(\gamma p+m)^{-1}$; this is sufficiently singular to yield infinite constants. It is shown below that near the cone, G' reduces to S_F'', the propagator corrected for gravitation only. Therefore it is sufficient to show that S_F'' is less singular than S_F. The complete problem, as represented by G', can then be expected to remain nonsingular, the omitted terms not being relevant in this connection. Alternately, the use of electrons and photons clothed by gravitation as the noninteracting elements of electrodynamics seems a sufficiently good approximation to bring out the finiteness brought into the problem by gravitation. Many-particle Green's functions are expected to display similar features at high energies, with the possible additional display of a bias against too close approach between the various particles (also to be expected on classical grounds). The self-effects constitute the main problem in any case. Gravitational field propagators can also be constructed, and should also behave nonsingularly; Schwarzschild-type singularities ought not to appear, since matter is now "regularly" distributed, while the "transverse" propagators should also be well behaved (since the gravitational self-interaction acts in the same way as it does on matter). We restrict ourselves to matter Green's functions here.

There should not appear any gravitational coupling renormalization, as the mass m plays the dual role of charge and (inertial) mass and the two types of renormalization, if any, must be equal by the equivalence of these two masses. Since there is only one type of "charge" here, no vacuum polarization is expected; the mass renormalization should also vanish in view of the absolute definition of mass and energy in the general theory, and of the fact that the observed mass is automatically defined from the coupling term between gravitation and matter, and through this by the value of P_0 for the system.

We first examine the usual Lorentz-covariant theory, to bring out the inevitable singular behavior of the Green's function of any field on the light cone, regardless of coupling. As a concrete example, we consider neutral scalar meson theory. Here the one-nucleon propagator, S_F', is given by[9]

$$S_F'(x-x')\equiv i\langle 0|\,(\psi(x)\bar{\psi}(x'))_+|0\rangle\epsilon(x,x')/\langle 0|0\rangle$$
$$=N^{-1}\int\psi(x)\bar{\psi}(x')\exp(iA[\psi,\varphi])\delta\varphi\delta\psi\delta\bar{\psi},\quad (2)$$

where the action A is

$$A=\int\bar{\psi}(x)(-i\gamma_\mu\partial_\mu+m)\psi(x)+g\int\bar{\psi}(x)\psi(x)\phi(x)$$
$$+\int\phi(x)(-\Box^2+\mu^2)\phi(x)\quad (3)$$

and the normalizing factor N corresponds to the vacuum-vacuum transition amplitude,

$$N=\int\exp(iA)\delta\varphi\delta\psi\delta\bar{\psi}.\quad (4)$$

ϵ is a sign factor, and the "+" sign a chronological ordering operator corresponding to outgoing wave boundary conditions. Integration over the nucleon variables can be performed, giving

$$S_F'(x-x')=N^{-1}\int G(x,x';\,\varphi)\exp[iA_M+A_0]\delta\varphi\quad (5)$$

$$[-i\gamma_\mu\partial_\mu+m+g\varphi(x)]G(x,x';\,\varphi)=\delta(x-x'),\quad (6)$$

where A_M is the free meson field part of A, A_0 an additional contribution due to closed loop effects; this term is a functional of φ but does not depend on the coordinates explicitly. The problem is thus resolved into two parts: obtaining the solution of Eq. (6) for G, and performing the integral in Eq. (5) with this result for G. The rigorous solution for G may be written[12]

$$G(x,x';\,\varphi)=S_F^{(0)}(x-x')F_1(x,x';\,\varphi)+F_2(xx';\,\varphi).\quad (7)$$

F_1 is a function everywhere regular in x,x' and approaching unity at the origin, $x=x'$. $S_F^{(0)}$ is the simple uncoupled Dirac propagator S_F with mass equal zero.[13] F_2 is holomorphic everywhere but on the light cone, $(x-x')^2=0$ where it may display a logarithmic singu-

[11] W. Pauli, Helv. Phys. Acta Suppl. 4, 69 (1956).

[12] The analysis used here of solutions of (6), (11), and the more general equation (14) was developed by J. Hadamard, *Lectures on Cauchy's Problem* (Dover Publications, New York, 1952). His work was improved by M. Riesz, Acta Math. 81, 1 (1949). The method of Riesz was adapted to the Dirac equation by H. Malmheden, Comm. Lund 8, 1 (1947). Since the relevant results are the same for our purposes, we deal mostly with the second-order equation for simplicity.

[13] The mass m can also be reinstated in S_F, the difference being absorbed in the definition of F.

larity; in some cases F_2 is null. The functional dependence on φ enters only in the F_i; further, we neglect F_2 here, since we are interested in the leading singularity on the light cone, and S_F is worse than logarithmic there.

Upon inserting (7) for G in the functional integral, we see that the singular $S_F(x-x')$ can be taken outside the integral. The integral of F, as $(x-x')^2\to 0$ may either converge absolutely in which case it represents a constant coefficient, the so-called Z_2 renormalization, on the cone, or, if it is illegitimate to interchange this limit with the integration, G' has an even worse singularity than S_F on the cone. At the origin G' is just equivalent to $S_F^{(0)}$, since $F_1=1$ there.[14] This theorem, proved earlier by other methods,[4] applies to a wide class of couplings, including all usual field theories. It expresses the fact that, as the arguments of any field's Green's function approach the light cone, the effects of its couplings can never result in smoother behavior than the free particle one, and probably cause a more singular variation.

In the parallel problem of interaction of a matter field (chosen here as scalar for simplicity) with the gravitational one, the total action is

$$A=\int(-g)^{\frac{1}{2}}[\varkappa^{-1}R+\tfrac{1}{2}\{-\phi\Box^2\phi+\mu^2\phi^2\}](d^4x),$$

where

$$\Box^2\equiv(-g)^{\frac{1}{2}}\partial_\mu(g^{\mu\nu}(-g)^{\frac{1}{2}}\partial_\nu), \tag{8}$$

where R is the scalar curvature. Interaction between the two fields arises through the metric dependence of the matter part of the action. Again we wish to compute

$$\Delta_F''(x,x')\equiv i\langle 0|\,(\phi(x)\phi(x'))_+|0\rangle/\langle 0|0\rangle$$

$$=N^{-1}\int\phi(x)\phi(x')\exp(iA)(\delta g)\delta\phi,$$

where

$$N=\int\exp(iA)(\delta g)\delta\phi. \tag{9}$$

The functional integration includes a summation over all possible Riemannian spaces preserving the signature $(+++-)$, the (δg) representing the appropriate variables with respect to which the integral is taken. We do not need their precise form; they might, for example, be the product of the 16 Vierbein variables at each point of space-time.[10] The arguments x, x' of Δ'' may at this stage be thought of as two quadruples of numbers representing two arbitrary points in a four-dimensional manifold, the integration over all possible Riemann spaces being an average of the two-point function over all possible ways of imposing a metric structure on this manifold, consonant with the signature and possibly certain topological requirements. Eliminating the matter variables by performing the φ integration (which corresponds to the ψ integration in the previous case), we get

$$\Delta_F''(x,x')=N^{-1}\int\Delta(x,x';\,g_{\mu\nu})\times\exp(iA_g+iA_{0g})(\delta g) \tag{10}$$

$$(-\Box^2+\mu^2)\Delta(x,x';\,g)=\delta(x,x') \tag{11}$$

where, in analogy to the previous case, A_g is the pure gravitational action, A_{0g} an addition to it due to the φ integral. Once again we must insert into (10) the rigorous solution Δ of (11). Here,

$$\Delta(x,x';\,g)=s_g^{-2}(x,x')F_1(x,x';\,g)+F_2(x,x';\,g). \tag{12}$$

s_g^{-2} is the inverse square of the geodesic distance between the two points x and x' in the space represented by the metric $g_{\mu\nu}(x)$ (there is a geodesic between two sufficiently close points in any Riemann space), F_1 is again a regular function everywhere, equal to unity at the origin, while F_2 has a logarithmic singularity on the light cone in the space (that is along all null geodesics issuing from x') if $\mu\neq 0$, and is again not of interest. The crucial difference between our previous examples and gravitation appears here: while the leading factor is, in strict analogy with the earlier problem, just the inverse square of the interval between x and x', it depends now on the dynamical variables of the metric field, and therefore cannot be removed outside the functional integral. There is no free particle propagation, since this factor also interacts with the metric. This is obviously true whatever the matter field is; all are coupled to $g_{\mu\nu}$. Had we considered G', that is, included other interactions as well, we would have obtained at this stage for the totally corrected propagator G',[15]

$$G'(x,x')=N^{-1}\int(\delta g)\exp(iA_g+iA_{0g})\times\int(\delta\chi)\exp[iA(\chi)]G(x,x';\,\chi,g), \tag{13}$$

where χ stands for all these additional interactions (electromagnetic, mesic, etc.) and

$$(-\Box^2+g_\mu[\chi]\partial_\mu+h[\chi])G(x,x';\,\chi,g)=\delta(xx'). \tag{14}$$

The χ dependence enters only through the g_μ and h, the coefficients of the second derivatives being the same as in (11), that is, only gravitational quantities. The

[14] The singularity of present field theory may be restated by the fact that in principle, at $x=x'$, $G'=S_F$, but that the limit $x\to x'$ yields an infinite proportionality constant between the two functions.

[15] The two-particle propagator would have another $G(y,y';\,g,\chi)$ in the integrand, the meson-nucleon one $\chi(\xi)\chi(\xi')$ there: in either case there appear to be no new singularity problems. It might even be expected that the distance between two-particles must exceed l_0.

general solution of (14), however, has its singularities determined entirely by the second derivative structure. That is why $s^{-2}(x,x')$ is always the leading term, and in usual theory can be moved out of the χ integral, but not past the gravitational integration. Incidentally, the equivalence of Δ'' and G' for purposes of investigating the effect of gravitation on the singular behavior is thereby established; at high enough energies all usual couplings become unimportant compared to the gravitational. Returning therefore to our result, (12), for $\Delta(x,x';g)$, we must now perform the functional integral, that is average over all spaces, with weight $\exp i(A_g+A_{0g})$ and measure (δg). We may again neglect F_1 and concentrate on s^{-2}, which has the same sort of singularity on the cone of the Riemann space that $(x-x')^{-2}$ has in Euclidian space. The average of this quantity, $\langle s^{-2}\rangle$, will give the leading behavior for all fields, since the details of other couplings, as well as masses, etc. are absorbed into the F's. Thus, if a cutoff appears (and l_0 is the only length now left) it will endow all particles with about the same "structure." Evaluation of the integral is not explicitly feasible. However, the following qualitative argument is in favor of its being nonsingular. Consider the two points x and x' to lie on each other's light cones in some given Riemann space, say, flat space. Then there exist, for any two such points, "almost" no other spaces for which this is simultaneously also true. "Almost" none means here that very stringent integrodifferential equations must be satisfied by the metrics of such other spaces. Further, the singularities of s^{-2} as a functional of the metric variables are not very strong. In analogy to the usual theory of integration, any measure in the function space which gives meaning to $\int(\delta g)$ will be such as to have no contributions from these "isolated" spaces. They will have measure zero. Conversely, if the two points are not joined by a null geodesic in flat space, there will be "almost" no other spaces in which they are so connected, and no new singularities will be brought in. The weight factor of the gravitational action, having magnitude unity, can only help in rendering the integration finite; it will also act (as the corresponding factor does in normal theory) to oscillate away the contributions from spaces very far away from the classical solutions, which make the phase large in a random way. A rigorous proof of our conjectures would seem to require a better developed mathematical theory of measure in functional integrations. Certain over-all properties of G' (or Δ'') can be obtained from corresponding ones of Δ; in particular, under interchange of x and x' and of boundary conditions, these Green's functions remain invariant; they are not, however, functions of $(x-x')$ since the spaces involved are not isotropic and homogeneous, unlike the flat one.

If we wish to consider effects near the light cone due to one of the ordinary interactions, whose contributions vanish at the cone, as we have seen, we may try to expand in powers of the interaction in question. For example, in the case of a scalar particle coupled to another, we have

$$G'=N^{-1}\int(\delta g)\exp(iA_g)\times\int\delta\varphi\,\exp(iA_\varphi)\Delta(x,x';\varphi,g)\qquad(15)$$

$$(-\Box^2+\mu^2+g\varphi)\Delta=\delta(x,x').\qquad(16)$$

Then we expand the F_i entering in the solution for Δ, and particularly F_1, since it comes in with the s^{-2} dominant term. Thus,

$$\Delta\sim s_g^{-2}(x,x')\Big[F_1(x,x';\varphi)\Big|_{\varphi=0}+g^2\int(dy)\times(dz)\varphi(y)\varphi(z)\delta^2F_1/\delta\varphi(y)\delta\varphi(z)\Big|_{\varphi=0}+\cdots\Big].\qquad(17)$$

The φ integration can be performed, and $\varphi(y)\varphi(z)\to\Delta(y,z;g)$. It may be thought that the infinities of perturbation theory (or indeed of local field theory in general, since its infinities first appear at this stage) will still arise in the integration over the intermediate points, because in any Riemann space (of the sum) overlapping singularities on the light cone still occur. However, the integration over the gravitational variables is to be performed first in order to include their averaging effect, before the intermediate integrals are done. Only then would the expansion be expected to make sense, since these apparent new singularities will have been smoothed out by the same argument as that given for $\langle s_g^{-2}\rangle$ above. Equivalently, the fluctuations of the matter field at a point may no longer be infinite, so that[16]

$$\lim_{x\to x'}\langle\varphi(x)\varphi(x')\rangle=\lim\langle s_g^{-2}(x,x')\rangle<\infty.\qquad(18)$$

In principle, the renormalization constants of, say, electrodynamics, are now presumably rendered finite by general relativity and can therefore be evaluated.

[16] J. Schwinger (unpublished lecture notes, Stanford 1956) relates the inconsistency of local field theory with the infinity arising at this stage, due to $\langle\varphi(x)\varphi(x')\rangle\to\infty$ as $x\to x'$. The gravitationally corrected theory is radically different from the usual one, so that care must be taken to use the now proper forms of such quantities as the physical mass or the equal time matter commutators (which now depend on the metric). As Schwinger has pointed out, the inconsistency of the usual theory results from the fact that it permits arbitrarily small space-time measurements, which, however, turn out to yield inconsistent (infinite) results because the interaction effects are always felt, however short the interval employed; while the equal time commutators require that no interaction effects be felt for sufficiently small times. When gravitation is taken into account, such arbitrarily short measurements cannot be made, as they would clash with the bounds established by general relativity; thus, the vacuum fluctuations of two-fields at a point or the overlap of light-cone singularities may no longer contribute to make results infinite for physical quantities.

The main effects of relativity, in fact, being high energy in nature, will enter into these constants, which absorb the asymptotic contributions in the theory. Asymptotic expansions of the functions of the theory could also be made, if everything were convergent, as this type of cutoff would be, by its very definition, physically acceptable. It is likely that the correct relation between the bare and physical charge and the cutoff depends strongly on the bare charge, which now, however, would be a finite parameter.[17] Of course, one could not attempt such expansions by transcribing directly into flat space-time language the form of the propagators obtained from general relativity, as they would then seem to violate some probability conservation or other physical requirements, since the very concept of cutoff in flat space-time violates them. Unfortunately, even the generally covariant form of the finite propagators seems difficult enough to extract from the functional integrals.[18]

The flat space theory is expected to emerge in the limit of vanishing gravitational coupling constant; however, since the latter appears in the total action as $\kappa^{-1}A_g$, this limit is quite different from the usual adiabatic shutting off procedure. In fact, $\kappa \to 0$ is equivalent to $\hbar \to 0$ in this term, that is, the classical limit; as is well known, the latter corresponds to having the phase stationary, in this case, $\delta A_g = 0$, i.e. the empty space field equations

$$R_{\mu\nu} = 0.$$

These equations become equivalent to the conditions for Euclidian space-time since we are dealing with closed systems, and asymptotically flat metrics accordingly. The $\int(\delta g)$ then acts to replace the metric wherever it occurs in the matter action or in the integrands by its Minkowski values. This decoupling procedure brings out the significance of the background flat space-time which serves as basis for classical measurements on the system. The flat space is thus analogous to the "in" and "out" quantities defined by means of the corresponding adiabatic decoupling in usual field theory, but in addition, the gravitational decoupling turns the metric quantities into c numbers.

IV. CONCLUSIONS

We have seen that a qualitative estimate of the gravitationally corrected matter propagators does indicate that singularities corresponding to point particles and interactions disappear. A corroborative argument, mentioned in Sec. I, comes from consideration of the energy-momentum vector, P_μ, and its commutation relations. In a flat space-time, and effectively (by suitable extension) in a nonflat but externally given geometry, Eq. (1) holds. P_μ determines the translation properties of the system, and because of the homogeneity and isotropy of the space, of any part of the system as well. In a space-time which interacts with matter, however, this relation ceases to be purely kinematical, and in fact no longer holds. P_μ does have, as a canonical conjugate, the center-of-mass coordinate, X_μ, of the system with respect to a flat background frame. However, P_μ cannot be used to locate closely any component of the system. The proof that any component field, if focused upon sufficiently, behaves as if uncoupled, which forms the basis of the difficulties in Lorentz-invariant theories, here becomes invalid, since it requires a P_μ with all the usual properties. It may be argued that the propagators themselves lose some of their meaning by the same argument, since they deal with localized effects of fields. This is true only to the extent that physical quantities always involve space-time integrals of such functions. The nonsingularity of the Δ'' and G' is an expression of the fact that one can never force such concentration of energy of a field that its gravitational couplings cease to matter—the induced geometry would cause a repulsion before such a stage could be reached.

Thus, a quantum field theory including gravitational effects seems to rid itself of point-singularity difficulties which neither part alone could avoid; classical relativity did tolerate point singularities in its solutions, which this more complete framework has no place for. Nonsingular matter distributions and nonsingular metric go hand in hand. This can be ascribed partly to the intimate coupling of geometry and kinetic energy, leading to limitations on energy concentrations, partly to the indefiniteness of the line element: the "smearing of the light cone" due to the appearance on an equal footing with Euclidian space of all other suitable Riemann spaces.

Even if the infinities are thus removed, the physical problems of obtaining correct dynamical descriptions of the interactions of observed particles (to say nothing of explaining their existence) are left untouched at this stage of the theory. However, it seems comforting that inclusion of the total, unmutilated general covariance avoids the formal difficulties to be found in Lorentz-invariant theory, and it may be that in the question of the values of the observed parameters, l_0 will play a significant role.

Since within the framework of flat space-time one could consider theories with interactions built into the kinetic energy terms, it is an interesting question whether finiteness can be achieved merely by such an

[17] L. D. Landau [*Niels Bohr and the Development of Physics* (Pergamon Press, London, 1955), p. 52] discusses an approximation to usual field theory which would imply a breakdown of the latter below a cut-off distance of about l_0.

[18] Practical evaluation of such functional integrals as Eq. (10) is beset by the difficulties that any tractable expansion loses some vital features of the full theory, such as nonlinearity, general covariance, self-interaction, etc. Thus the linear approximation to the gravitational field corresponds to reinstatement of the privileged character of the Minkowski metric, since gravitation now propagates like any linear field in flat space-time. Further, this approximation on the metric entering in the matter part is probably invalid just at the high energies of interest. It is hoped to treat some of the approximation methods in a future publication.

extension, or whether the free parts of those fields which play such a role need, in addition, to be nonlinear in some specific way, or whether only general relativistic coupling will work. Equation (1) would still hold for such extended theories, being a purely kinematical property of homogeneous space-time, but the usual local commutativity properties would be lost; that is, the commutator of a field component with its time-derivative would, even at equal times, depend on other fields. Alternately, the conjugate momentum would be a function of such other fields as well as of the time derivative of the corresponding component. The assumption that this is not the case appears to be implicitly required in the proofs of the pessimistic theorems. The commutation relations are closely related to the creation and annihilation of particles; if they now depend on the other fields present, it might happen that the contributions, as the thresholds of higher and higher creation processes are passed with increasing energies, are damped thereby. These contributions from new creation processes seem to be the cause (or another expression) of the divergences in field theory, in which cases the new couplings might yield convergence. Such couplings arise, for example, in quantum hydrodynamics.[19] Saturation might be expected to occur with some of these couplings; that is, the presence of many quanta, or high energies, may damp further creation.[20]

ACKNOWLEDGMENTS

I am most thankful to Professor O. Klein for many valuable discussions of this work; his contributions to all its phases have been extremely illuminating.

The author also wishes to acknowledge the hospitality accorded him at the Institute for Theoretical Physics. His work was performed during tenure of a National Science Foundation Fellowship.

[19] A. Thellung, Helv. Phys. Acta **29**, 103 (1956). In this case the kinetic energy has the form $\rho v^2/2$, both ρ and v being fields. This form may be very different, however, from one in which several variables ρ_{ij} multiply v_iv_j.

[20] Landau[17] has also remarked that at high energies field theory might go over into a quantum hydrodynamical scheme.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Negative Mass in General Relativity

H. Bondi

King's College, University of London, London, England

1.

Although some of the arguments relating to negative mass are fairly elementary and well known, it will nevertheless be of advantage to rediscuss the meaning of this term. In the first instance, without fully specifying a theory, we can distinguish between three kinds of mass according to the measurement by which it is defined: inertial, passive gravitational, and active gravitational mass. Inertial mass is the quantity that enters (and is defined by) Newton's second law*; passive gravitational mass is the mass on which the gravitational fields acts, that is it is defined by $\mathbf{F}=-m\,\mathrm{grad}V$; active gravitational mass is the mass that is the source of gravitational fields and is hence the mass that enters Poisson's equation and Gauss' law.

In Newtonian physics the law of action and reaction implies the equality of active and passive gravitational masses, but the equality of inertial mass with these other two is a separate empirical fact. The sign of both these masses can take either value and it is an additional empirical result that it is always positive. Four cases accordingly arise, if this empirical fact is left out of account.

(i) All mass is positive; this is familiar.

(ii) Inertial mass negative, gravitational mass is positive. A body consisting of matter of this kind will respond perversely to all forces whether gravitational or of other kinds, but will produce gravitational forces just as a usual body does.

(iii) Inertial mass positive, gravitational masses negative. In this case we would have normal behavior relating to all nongravitational forces, but gravitational behavior involving masses of this type and of type (i) would be governed by a negative Coulomb law; i.e., like masses would attract and unlike masses would repel.

(iv) All mass is negative. This would be a combination of (ii) and (iii). Matter of this kind responds perversely to nongravitational forces, responds like ordinary matter to gravitational forces, but produces repulsive gravitational fields.

In general relativity the situation is quite different. The principle of equivalence is not a separate fact but is basic to the theory. Accordingly the ratio of inertial and passive gravitational masses is the same for all bodies. The relation between active and passive gravitational masses is not fixed by anything like Newton's third law as this would require integrals over

* A mass-independent force (say, of electromagnetic nature) has to be used here, for obvious reasons.

extended regions of space-time which do not possess the required tensorial character. A good deal of work has been done[1,2] and tends to indicate that the relation is rather complicated.

As long as relativity is considered purely as a theory of gravitation, the inertial and passive gravitational masses do not in fact appear. Active gravitational mass occurs for the first time as a constant of integration in Schwarzschild's solution. If this constant is taken to be positive, then test particles will, in the first approximation, describe the Newtonian orbits corresponding to an attractive body. If, however, the constant is taken to be negative then, in the first approximation, test particles will describe the orbits corresponding to the Newtonian case with repulsion. Note that in the first case *all* bodies will be attracted, in the second *all* bodies will be repelled.

If we now leave the one-body problem and consider the two-body case, then a remarkable situation arises. Imagine a body of positive mass and a body of negative mass separated by empty space. Then, to use the language of the Newtonian approximation, the positive body will attract the negative one (since all bodies are attracted by it), while the negative body will repel the positive body (since all bodies are repelled by it). If the motion is confined to the line of centers, then one would expect the pair to move off with uniform acceleration. This rather surprising result clearly requires confirmation by the complete construction of the model in general relativity.

2.

Uniformly accelerated systems in general relativity (and in special relativity) are well known,[3] but it may be worth while repeating here briefly the main properties of such systems. The Newtonian concept of uniform acceleration may be generalized to special relativity in a number of ways, but one of these is of outstanding importance in retaining the stationary property. The system in this case is described by the equation $\zeta^2-\tau^2=\text{const}$. The orbits of all particles in the $\tau-\zeta$ plane form a system of rectangular hyperbolas with fixed asymptotes. These particles all have an acceleration which is uniform in the sense that the motion of each particle viewed from that particle is constant in time. However, although the acceleration of every point is uniform, the acceleration of different points is not the same. Roughly speaking, the nearer the trajectory of the particle passes to the origin the larger its acceleration. The remarkable feature of this system is its stationary character. If any particle carried an observer measuring the distance of any other particle partaking of the motion, then he would find this distance to be constant in time, although to a fixed observer he and the other particle would have different accelerations. The usual Minkowski metric may be transformed to the uniformly accelerated frame by the transformation $\tau=z\sinh t$, $\zeta=z\cosh t$, $\xi=x$, $\eta=y$, resulting in the metric $ds^2=z^2dt^2-dx^2-dy^2-dz^2$. It is clear that this metric does not cover the whole of space, but is bordered by portions of the asymptotes which act as horizons and accordingly the metric only covers one-quarter of space-time (Fig. 1).

We now return to the task of constructing a model in general relativity of two bodies whose masses have opposite sign. Accordingly we use a uniformly accelerated frame, and then immerse in it two finite bodies with opposite sign of mass. In the uniformly accelerated frame the system will be axially symmetric and so we can use[4] the metric of Weyl and Levi-Civita. In this metric we have, in empty space,

$$ds^2=e^{2\varphi}d\bar{t}^2-e^{-2\varphi}[e^{2\sigma}(d\bar{r}^2+d\bar{z}^2)+\bar{r}^2d\theta^2] \tag{1}$$

where $\varphi=\varphi(\bar{r},\bar{z})$, $\sigma=\sigma(\bar{r},\bar{z})$ satisfy

$$\left[\frac{\partial^2}{\partial\bar{r}^2}+\frac{1}{\bar{r}}\frac{\partial}{\partial\bar{r}}+\frac{\partial^2}{\partial\bar{z}^2}\right]\varphi=0. \tag{2}$$

[The operator occurring in (2) will be denoted by ∇^2.]

$$\frac{\partial\sigma}{\partial\bar{r}}=\bar{r}\left[\left(\frac{\partial\varphi}{\partial\bar{r}}\right)^2-\left(\frac{\partial\varphi}{\partial\bar{z}}\right)^2\right] \tag{3}$$

$$\frac{\partial\sigma}{\partial\bar{z}}=2\bar{r}\frac{\partial\varphi}{\partial\bar{r}}\frac{\partial\varphi}{\partial\bar{z}}. \tag{4}$$

It is well known that there is a consistency condition for this metric.[4] For our purposes this condition will be given in a slightly different form from the usual one.

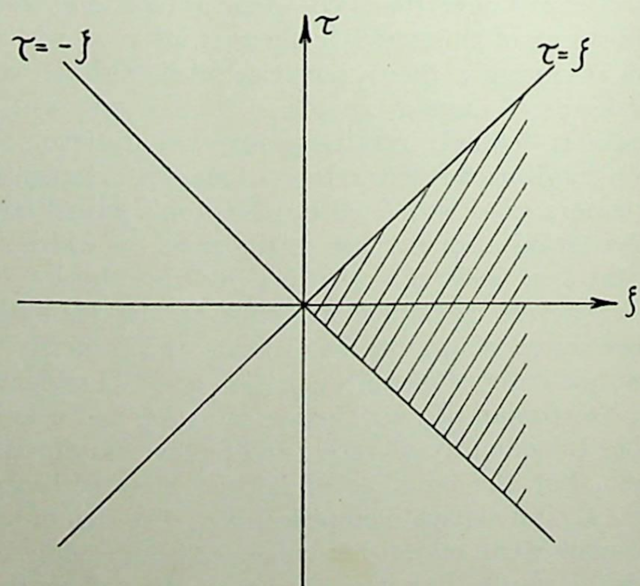


FIG. 1. The shaded portion on the right is mapped on the half-space $z\geqslant 0$ and also on the entire $(\bar{t},\bar{r},\bar{z},\theta)$ space.

[1] J. L. Synge, Proc. Edinburgh Math. Soc. (2) 7, 93 (1937).
[2] E. T. Whittaker, Proc. Roy. Soc. (London) A149, 384 (1935).
[3] L. Marder, Proc. Cambridge Phil. Soc. 53, 194 (1957).
[4] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice Hall, Inc., New York, 1946), p. 208.

For any closed circuit C situated entirely in empty space (though possibly enclosing a nonempty region) we have from (3) and (4)

$$0=\oint_C\left[\frac{\partial\sigma}{\partial\bar{r}}d\bar{r}+\frac{\partial\sigma}{\partial\bar{z}}d\bar{z}\right]=\oint_C\left\{\bar{r}\left[\left(\frac{\partial\varphi}{\partial\bar{r}}\right)^2-\left(\frac{\partial\varphi}{\partial\bar{z}}\right)^2\right]d\bar{r}+2\bar{r}\frac{\partial\varphi}{\partial\bar{r}}\frac{\partial\varphi}{\partial\bar{z}}d\bar{z}\right\}=2\int\frac{\partial\varphi}{\partial\bar{z}}\nabla^2\varphi\,\bar{r}d\bar{r}d\bar{z} \quad (5)$$

by the divergence theorem, the integral being taken through the part of the meridian plane enclosed by C. It has been assumed here that through the whole of this region the metric has the form (1) though it need not satisfy Eqs. (2), (3), and (4).

The metric (1) is not in fact the most general metric for a nonempty static axially symmetric region but it is sufficiently general for the model we wish to construct. It is easy to show that

$$-\kappa\rho=-\kappa T_0{}^0=e^{2(\varphi-\sigma)}\left[-2\nabla^2\varphi+\nabla^2\sigma-\frac{1}{\bar{r}}\frac{\partial\sigma}{\partial\bar{r}}+\left(\frac{\partial\varphi}{\partial\bar{r}}\right)^2+\left(\frac{\partial\varphi}{\partial\bar{z}}\right)^2\right] \quad (6)$$

$$-\kappa p_{11}=-\kappa T_1{}^1=\kappa T_2{}^2=\kappa p_{22}=e^{2(\varphi-\sigma)}\left[\frac{1}{\bar{r}}\frac{\partial\sigma}{\partial\bar{r}}-\left(\frac{\partial\varphi}{\partial\bar{r}}\right)^2+\left(\frac{\partial\varphi}{\partial\bar{z}}\right)^2\right] \quad (7)$$

$$-\kappa p_{33}=-\kappa T_3{}^3=e^{2(\varphi-\sigma)}\left[\nabla^2\sigma-\frac{1}{\bar{r}}\frac{\partial\sigma}{\partial\bar{r}}+\left(\frac{\partial\varphi}{\partial\bar{r}}\right)^2+\left(\frac{\partial\varphi}{\partial\bar{z}}\right)^2\right] \quad (8)$$

$$-\kappa T_{12}=2\frac{\partial\varphi}{\partial\bar{r}}\frac{\partial\varphi}{\partial\bar{z}}-\frac{1}{\bar{r}}\frac{\partial\sigma}{\partial\bar{z}}. \quad (9)$$

The type of specialization involved in the retention of metric (1) is clearly shown in (7). The condition that T_{12} is finite on the axis implies that $\sigma=\text{const}$ on $\bar{r}=0$. Without loss of generality we may take $\sigma=0$ on $\bar{r}=0$. If we suppose that there is no matter on $\bar{r}=0$, then this condition also follows from (5).

We can now construct a Newtonian analog of our system in which $\bar{r}$, $\bar{z}$, θ are cylindrical polar coordinates and in which φ is the gravitational potential (in gravitational units). In empty space φ satisfies Laplace's equation (2). Also, as long as φ is small, (3) and (4) imply that σ is small of the second order. Equation (6) is then, to the first order, identical with Poisson's equation, while Eqs. (7), (8), and (9) imply that the stresses are small compared with the density.

The significance of the consistency condition (5) is now clear: it is that the gravitational force on any body surrounded by empty space must have a vanishing $\bar{z}$ component. Since the other components vanish by symmetry, (5) is simply the Newtonian equilibrium condition. We can now regard φ as the *exact* Newtonian potential of a Newtonian analog system though the density of the Newtonian system will not be exactly the same as the density of the relativistic system.

Since Laplace's equation is linear, we can of course superpose solutions. In particular if there are two bodies [i.e., two separate regions in which (2) does not hold], then the two corresponding solutions may be superposed, subject to condition (5) holding for each body separately. A few theorems of Newtonian gravitation may now usefully be quoted.

(i) If $\varphi\rightarrow 0$ at ∞ and $\nabla^2\varphi=0$ except in a single finite closed region, then condition (5) is satisfied.

(ii) If $\varphi\rightarrow 0$ at ∞ and $\nabla^2\varphi=0$ except in a finite closed region which lies entirely in the region $\bar{z}<a$ and in which $\nabla^2\varphi\geqslant 0$, then $\partial\varphi/\partial\bar{z}\geqslant 0$ for all $\bar{z}\geqslant a$.

(iii) If $\varphi\rightarrow 0$ at ∞ and $\nabla^2\varphi=0$ except in two finite regions one of which lies entirely in $\bar{z}<a$ and the other in $\bar{z}>a$, and in each of which $\nabla^2\varphi$ is of one sign, then condition (5) cannot be satisfied. This important result follows from (i) and (ii). If φ is split into two parts, φ_1 satisfying Laplace's equation except in body 1 and φ_2 except in body 2, then condition (5) for body 1 requires that

$$\iint\left[\frac{\partial\varphi_1}{\partial\bar{z}}\nabla^2\varphi_1+\frac{\partial\varphi_2}{\partial\bar{z}}\nabla^2\varphi_1\right]\bar{r}d\bar{r}d\bar{z}=0.$$

The first term vanishes by (i); the second cannot vanish by virtue of (ii). The theorem shows that there is no static solution in general relativity for two bodies, each containing matter of one sign, situated on opposite sides of a surface $z=\text{const}$, with the metric tending to the Minkowski metric at infinity. It will now be shown that if the last condition is dropped such a solution is no longer impossible.

3.

The uniformly accelerated metric

$$ds^2=z^2dt^2-dx^2-dy^2-dz^2 \quad (10)$$

may be transformed to the Weyl-Levi-Civita form (1) by the transformation

$$\left.\begin{aligned} t&=\bar{t}\\ z&=e^{\varphi}\\ (x^2+y^2)^{\frac{1}{2}}&=\bar{r}e^{-\varphi}\\ \tan^{-1}y/x&=\theta\end{aligned}\right\} \quad (11)$$

where

$$\varphi=\tfrac{1}{2}\log\{[\bar{r}^2+(\bar{z}-a)^2]^{\frac{1}{2}}+(\bar{z}-a)\} \quad (12)$$

$$\sigma=\tfrac{1}{2}\log\{\tfrac{1}{2}+\tfrac{1}{2}(\bar{z}-a)[\bar{r}^2+(\bar{z}-a)^2]^{-\frac{1}{2}}\} \quad (13)$$

and a is an arbitrary constant.

This φ, which will be called φ_0, satisfies Laplace's equation in the whole relevant portion of space, i.e., except on $\bar{r}=0$, $\bar{z}\leqslant a$. As was discussed earlier, the uniformly accelerated metric only represents part of space-time and so the singularity at $\bar{r}=0$, $\bar{z}\leqslant a$ need not cause surprise or alarm.

Note that, for $\bar{z}>a$, $\partial\varphi_0/\partial\bar{z}>0$.

Consider now the problem of two bodies, as in (iii) above, but add the potential φ_0. This implies that we drop the boundary condition that space is Minkowskian at infinity and replace it by that appropriate to a uniformly accelerated frame of reference in a space-time which is flat at infinity. If each of the bodies is entirely in $\bar{z}>a$, and if in the body (body 1) in the lower $\bar{z}$ region $\nabla^2\varphi_1\leqslant 0$, whereas in the other one (body 2) $\nabla^2\varphi_2\geqslant 0$, then condition (5) can be satisfied in suitable circumstances. For then, inside body 1, $\partial\varphi_2/\partial\bar{z}<0$ and, in body 2, $\partial\varphi_1/\partial\bar{z}<0$. As $\partial\varphi_0/\partial\bar{z}>0$ in both regions, there is now no argument from signs to show that the arrangement is impossible.

To establish the possibility we proceed as follows. Let B_1, B_2 be two finite regions of space with B_1 entirely in $\bar{z}_1\leqslant\bar{z}\leqslant\bar{z}_1'$ and B_2 in $\bar{z}_2\leqslant\bar{z}\leqslant\bar{z}_2'$ where $\bar{z}_1'<\bar{z}_2$. Let φ_1 satisfy Laplace's equation everwhere outside B_1, with $\varphi_1\rightarrow 0$ at infinity and with $\nabla^2\varphi_1\leqslant 0$ in B_1. Similarly let φ_2 satisfy Laplace's equation outside B_2, let it tend to zero at infinity, but $\nabla^2\varphi_2\geqslant 0$ in B_2. Then consider

$$\varphi=\varphi_0+k\varphi_1+l\varphi_2 \tag{14}$$

where the a entering φ_0 is a constant and k, l are constants to be determined later, with l, $k>0$ and $a<\bar{z}_1$. If condition (5) is to be satisfied for both B_1 and B_2 then

$$k\int\int_{B_1}\bar{r}d\bar{r}d\bar{z}\nabla^2\varphi_1\left[\frac{\partial\varphi_0}{\partial\bar{z}}+l\frac{\partial\varphi_2}{\partial\bar{z}}\right]=0 \qquad (15)\dagger$$

and

$$l\int\int_{B_2}\bar{r}d\bar{r}d\bar{z}\nabla^2\varphi_2\left[\frac{\partial\varphi_0}{\partial\bar{z}}+k\frac{\partial\varphi_1}{\partial\bar{z}}\right]=0. \qquad (16)\dagger$$

Since the factor k can be canceled in (15), this equation can be considered as an equation determining l. Moreover, since in B_1

$$\frac{\partial\varphi_0}{\partial\bar{z}}>0>\frac{\partial\varphi_2}{\partial\bar{z}}, \tag{17}$$

l will be positive. Similarly (16) determines a positive k.

All the conditions of the problem, therefore, are satisfied and so we have succeeded in constructing a uniformly accelerated pair of bodies whose densities have opposite sign.

† The Newtonian "self-force" term $\nabla^2\varphi_s\partial\varphi_s/\partial\bar{z}$ has been omitted in (15) and (16), since its integral vanishes by theorem (i).

As an example we may take $a=0$ and

$$\varphi_1=\begin{cases}-\dfrac{m_1}{[\bar{r}^2+(\bar{z}-h_1)^2]^{\frac{1}{2}}} & (\bar{r}^2+(\bar{z}-h_1)^2\geqslant a_1^2)\\[2ex] \dfrac{m_1}{2a_1^3}[\bar{r}^2+(\bar{z}-h_1)^2]-\dfrac{3m_1}{2a_1} & (\bar{r}^2+(\bar{z}-h_1)^2\leqslant a_1^2)\end{cases} \tag{18}$$

$$\varphi_2=\begin{cases}-\dfrac{m_2}{[\bar{r}^2+(\bar{z}-h_2)^2]^{\frac{1}{2}}} & (\bar{r}^2+(\bar{z}-h_2)^2\geqslant a_2^2)\\[2ex] \dfrac{m_2}{2a_2^3}[\bar{r}^2+(\bar{z}-h_2)^2]-\dfrac{3m_2}{2a_2} & (\bar{r}^2+(\bar{z}-h_2)^2\leqslant a_2^2)\end{cases}. \tag{19}$$

Here, for simplicity, h_1 and h_2 are disposable rather than k and l.

Condition (16) becomes, since $\nabla^2\varphi_2=\text{const}$ in B_2,

$$0=\int\int_{B_2}\bar{r}d\bar{r}d\bar{z}\left[\frac{\partial\varphi_0}{\partial\bar{z}}+\frac{\partial\varphi_1}{\partial\bar{z}}\right]. \tag{20}$$

If $a_2\ll h_2-h_1$ then this becomes effectively

$$\left[\frac{\partial\varphi_0}{\partial\bar{z}}+\frac{\partial\varphi_1}{\partial\bar{z}}\right]_{\bar{z}=h_2,\bar{r}=0}=0 \tag{21}$$

and so

$$\frac{1}{2h_2}=-\frac{m_1}{(h_2-h_1)^2}. \tag{22}$$

Hence m_1 is negative.

Similarly we must have, if $a_1\ll h_2-h_1$ and $a_1\ll h_1$,

$$\left(\frac{\partial\varphi_0}{\partial\bar{z}}+\frac{\partial\varphi_2}{\partial\bar{z}}\right)_{\bar{z}=h_1,\bar{r}=0}=0 \tag{23}$$

$$\frac{1}{2h_1}=\frac{m_2}{(h_2-h_1)^2}. \tag{24}$$

It follows that in both bodies the derivatives of φ_0 are of the same order of magnitude as the derivatives of the potentials produced by themselves. If, then, the Newtonian potentials of both bodies are small, the derivatives of σ will be small of the second order, in spite of the additional terms due to φ_0. Accordingly the densities are an order of magnitude larger than the stresses, and are themselves, to the first order, given by $\kappa\rho=2e^{2(\varphi_0-\sigma_0)}\nabla^2\varphi$.

By (12) and (13)

$$e^{-2(\varphi_0-\sigma_0)}=\frac{1}{2(\bar{r}^2+\bar{z}^2)^{\frac{1}{2}}}\doteq\frac{1}{2h}. \tag{25}$$

Hence

$$\kappa\rho=12h\frac{m}{a^3}. \tag{26}$$

It may be advantageous to view the system from the Galilean frame of reference τ, ζ, α, θ at $\tau=\bar{t}=0$.‡ Then the bodies appear to be spheres with centers at $\zeta=(2h)^{\frac{1}{2}}$ and of radii $a(2h)^{-\frac{1}{2}}$. The densities are still given by (26) and so the masses are $M=m(2h)^{-\frac{1}{2}}$, while the accelerations are $(2h)^{-\frac{1}{2}}$. In the Newtonian limit, the accelerations should be given by

$$\frac{1}{(2h_2)^{\frac{1}{2}}}=f_2=-\frac{M_1}{[(2h_2)^{\frac{1}{2}}-(2h_1)^{\frac{1}{2}}]^2}$$

$$=-\frac{m_1}{(2h_1)^{\frac{1}{2}}[(2h_2)^{\frac{1}{2}}-(2h_1)^{\frac{1}{2}}]^2}. \quad (27)$$

By (22) this equals

$$\frac{(h_2-h_1)^2}{2h_2(2h_1)^{\frac{1}{2}}[(2h_2)^{\frac{1}{2}}-(2h_1)^{\frac{1}{2}}]^2}$$

$$=\frac{1}{(2h_2)^{\frac{1}{2}}}\frac{\{\frac{1}{2}[(2h_1)^{\frac{1}{2}}+(2h_2)^{\frac{1}{2}}]\}^2}{(2h_1)^{\frac{1}{2}}(2h_2)^{\frac{1}{2}}}. \quad (28)$$

This will be approximately the case provided $h_2-h_1\ll h_1$; i.e., provided the product of acceleration and distance apart is small. This is an appropriate limitation for the Newtonian case. It is interesting to note that the masses are not quite equal and opposite, but this is not surprising since their accelerations must be unequal in the uniformly accelerated model.

4.

The metric constructed in the preceding section contains a singularity at $\bar{r}=0$, $\bar{z}\leqslant 0$. This is not surprising, since the uniformly accelerated metric $\varphi=\varphi_0$ also contains such a singularity which is, however, purely artificial and is transformed away by returning to the $(\tau,\zeta,\alpha,\theta)$ metric. Is a similar elimination of the singularity possible if $\varphi=\varphi_0+\varphi_1+\varphi_2$?

We note first of all that if $\varphi=\varphi_0$ the entire $(\bar{t},\bar{z},\bar{r},\theta)$ space corresponds to $z\geqslant 0$ in (t,z,r,θ) and to a quarter of space-time; *viz.*, $\zeta\geqslant|\tau|$ in $(\tau,\zeta,\alpha,\theta)$. Keeping to the same transformation equations

$$\bar{r}=rz,\quad \bar{z}=\tfrac{1}{2}(z^2-r^2),\quad \bar{t}=t \quad (29)$$

we have now the metric

$$ds^2=z^2e^{2\psi}dt^2-e^{2(\delta-\psi)}(dr^2+dr^2)-r^2e^{-2\psi}d\theta^2 \quad (30)$$

where the singular part of the metric appears explicitly and we are only dealing with $z\geqslant 0$. It is easily seen that

$$\psi(z,r)=\varphi(\bar{z},\bar{r})-\varphi_0(\bar{z},\bar{r}) \quad (31)$$

is a regular function of z and r. Moreover it is readily established that, for small z, ψ can be expanded as a power series in z^2 with coefficients depending on r, the coefficients themselves being representable by power series in r^2 for small r. Also

$$\delta(z,r)=\sigma(\bar{z},\bar{r})-\sigma_0(\bar{z},\bar{r}) \quad (32)$$

is of the same nature.

Defining now

$$\tau=ze^{\psi}\sinh t,\quad \zeta=ze^{\psi}\cosh t,\quad \alpha=re^{-\psi} \quad (33)$$

we arrive at the metric

$$ds^2=d\tau^2-d\zeta^2-\mu(\tau d\tau-\zeta d\zeta)^2-2\nu d\alpha(\tau d\tau-\zeta d\zeta)$$
$$-\lambda d\alpha^2-\alpha^2d\theta^2 \quad (34)$$

where the coefficients μ, ν, λ are functions only of α and $\zeta^2-\tau^2$. Note that this metric is invariant under any (ζ,τ) Lorentz transformation, showing that we are still dealing with a case of uniform acceleration though we are no longer in flat space-time.

By virtue of Eqs. (33), metric (34) is established only for $\zeta\geqslant|\tau|$. What happens on and beyond this boundary? It is clear from the structure of (34) that the boundary consists of parts of two null geodesics. Furthermore, a somewhat laborious comparison of coefficients yields expressions for μ, ν, λ in terms of ψ, $\partial\psi/\partial z$, $\partial\psi/\partial r$, and δ. An examination of these expressions establishes that, as a consequence of the behavior of ψ and δ for small z referred to above, the three new coefficients can be expanded in series of powers of $\zeta^2-\tau^2$ near the boundary, the coefficients of the terms in these series being functions of α that themselves can be expanded in powers of α^2 near $\alpha=0$. Metric (34) is therefore perfectly regular at the boundary.

As the boundary is a null geodesic the continuation of the metric beyond it is not uniquely defined. It would be most attractive to find a continuation of (34) that was free of singularities and of matter, and comprised all space-time. The mathematical difficulties of finding such a continuation appear to be formidable and have so far proved unsurmountable. It is also an interesting and significant problem to establish the existence of such a continuation, but this too has so far defied solution.

A solution of a different character has however been obtained. If (34) is supposed to retain its form for all τ and ζ then a metric is obtained that is symmetrical about $\tau=0$ and about $\zeta=0$. In the region $\tau\geqslant|\zeta|$ the coefficients depend only on α and on the *time like* variable $\tau^2-\zeta^2$ (and similarly in $|\zeta|\leqslant-\tau$) while in $-\zeta\geqslant|\tau|$ the mirror image of the conditions in $\zeta\geqslant|\tau|$ will apply; i.e., there will again be two uniformly accelerated bodies of opposite sign of mass. The sole question to be decided in order to establish the validity of this type of solution is whether there is an empty space metric of this kind in $\tau\geqslant|\zeta|$ fitting with the required degree of smoothness on to our previously obtained metric at $\tau=\zeta\geqslant 0$.

We first notice that the transformation

$$\tau=Te^{\epsilon}\cosh Z,\quad \zeta=Te^{\epsilon}\sinh Z,\quad \alpha=Re^{-\epsilon}$$

‡ Where $\xi=\alpha\cos\theta$, $\eta=\alpha\sin\theta$.

with

$$\epsilon=\epsilon(T,R) \tag{35}$$

yields the metric

$$ds^2=e^{2(\eta-\epsilon)}(dT^2-dR^2)-T^2e^{2\epsilon}dZ^2-R^2e^{-2\epsilon}d\theta^2 \tag{36}$$

with

$$\eta=\eta(T,R)$$

provided the μ, ν, λ of (34) are suitably connected with the ϵ, η of (36). In fact it turns out that the expressions for μ, ν, λ in terms of ϵ and η are identical with those in terms of ψ and δ provided ψ, δ, r, z^2, $z\partial/\partial z$ are replaced respectively by ϵ, η, R, $-T^2$, $T\partial/\partial T$. Accordingly, we have found a metric fitting smoothly to our original metric by continuing the intermediate metric (34) across the boundary.

The behavior of (36) away from the boundary is governed in empty space by the equations

$$\frac{\partial^2\epsilon}{\partial R^2}+\frac{1}{R}\frac{\partial\epsilon}{\partial R}=\frac{\partial^2\epsilon}{\partial T^2}+\frac{1}{T}\frac{\partial\epsilon}{\partial T} \tag{37}$$

$$(R^2-T^2)\frac{\partial\eta}{\partial T}=R^2T\left[\left(\frac{\partial\epsilon}{\partial R}\right)^2+\left(\frac{\partial\epsilon}{\partial T}\right)^2\right]-2RT^2\frac{\partial\epsilon}{\partial R}\frac{\partial\epsilon}{\partial T}+2R^2\frac{\partial\epsilon}{\partial T}-2RT\frac{\partial\epsilon}{\partial R} \tag{38}$$

$$-(R^2-T^2)\frac{\partial\eta}{\partial R}=RT^2\left[\left(\frac{\partial\epsilon}{\partial R}\right)^2+\left(\frac{\partial\epsilon}{\partial T}\right)^2\right]-2R^2T\frac{\partial\epsilon}{\partial R}\frac{\partial\epsilon}{\partial T}+2RT\frac{\partial\epsilon}{\partial T}-2R^2\frac{\partial\epsilon}{\partial R}. \tag{39}$$

Since (37) is a hyperbolic equation it will have no singularities for $T>0$, $R\geqslant 0$ provided ϵ and $\partial\epsilon/\partial T$ are given on $T=0$ in a nonsingular manner as even functions of R. This is the case in view of the properties of ψ and since ϵ is connected with ψ as mentioned above. In fact, near $T=0$, ϵ can be expanded in a power series in T^2 with coefficients that themselves can be expanded in powers of R^2, as a consequence of the corresponding property of ψ. Accordingly ϵ will be nonsingular. As for η, the only doubt about its nonsingular character arises from the factor R^2-T^2. It is, however, easily seen that this factor does not lead to a singularity provided $\partial\epsilon/\partial R=\partial\epsilon/\partial T$ on $R=T$, and this in turn is a consequence of (37) and of the fact that $\partial\epsilon/\partial R=\partial\epsilon/\partial T=0$ at $R=T=0$.

We have succeeded, therefore, in constructing a world-wide nonsingular solution of Einstein's equations containing two oppositely accelerated pairs of bodies,§ each pair consisting of two bodies of opposite sign of mass. Since $T_0{}^0$ and m are for any one body of the same sign the negative mass occurring is of type (iv).

§ This solution is closely analogous to Born's solution for the electromagnetic case [H. Bondi and T. Gold, Proc. Roy. Soc. (London) A229, 416 (1955)].

Gravitational Field of an Axially Symmetric System in First Approximation

NATHAN ROSEN AND HADASSAH SHAMIR

Israel Institute of Technology, Haifa, Israel

THE question as to whether a material system can radiate gravitational waves, on the basis of the field equations of the general theory of relativity, does not appear to have been answered conclusively. The present work is intended as a first step in looking for the answer.

To investigate the possibility of a physical system radiating gravitational waves it is desirable to choose a simple system, one with axial symmetry. If the field of such a system is described by means of a spherical polar coordinate system $(x^1,x^2,x^3,x^4)\equiv(r,\vartheta,\varphi,t)$,then by a suitable choice of coordinates one can satisfy two conditions: (a) the metric tensor $g_{\mu\nu}$ is independent of the angle φ; (b) it is diagonal.

The field equations in the empty space surrounding the system,[1]

$$G_{\mu\nu}\equiv R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=0, \qquad (1)$$

give a set of 7 equations, since $G_{\mu\nu}$ vanishes identically if $(\mu,\nu)=(1,3),\ (2,3),\ (3,4)$. These equations serve to determine the 4 diagonal components of $g_{\mu\nu}$. Among the equations there exist 3 identities, the contracted Bianchi identities, except for the one with index 3, which is trivial.

The field equations are nonlinear and difficult to solve exactly. It is proposed to solve them by the method of successive approximations. As a beginning, the first approximation is obtained here. Let us write the line element in the form

$$ds^2=-(1+\rho)dr^2-r^2(1+\sigma)d\vartheta^2 -r^2\sin^2\vartheta(1+\tau)d\varphi^2+(1+\mu)dt^2, \qquad (2)$$

where ρ, σ, τ, and μ are functions of r, ϑ, and t and are considered (along with their derivatives) to be small of the first order. The linear approximation of the field equations has the following form (indexes denoting partial differentiation):

$$\sigma_{44}+\tau_{44}-\frac{1}{r^2}(\tau_{22}+\mu_{22})+\frac{\text{ctg}\vartheta}{r^2}(\sigma_2-2\tau_2-\mu_2) -\frac{1}{r}(\sigma_1+\tau_1+2\mu_1)+\frac{2}{r^2}(\rho-\sigma)=0, \qquad (3a)$$

$$-\tau_{11}-\mu_{11}+\rho_{44}+\tau_{44}-\frac{\text{ctg}\vartheta}{r^2}(\rho_2+\mu_2) +\frac{1}{r}(\rho_1-2\tau_1-\mu_1)=0, \qquad (3b)$$

$$-\sigma_{11}-\mu_{11}+\rho_{44}+\sigma_{44}-\frac{1}{r^2}(\rho_{22}+\mu_{22}) +\frac{1}{r}(\rho_1-2\sigma_1-\mu_1)=0, \qquad (3c)$$

$$\sigma_{11}+\tau_{11}+\frac{1}{r^2}(\rho_{22}+\tau_{22})+\frac{\text{ctg}\vartheta}{r^2}(\rho_2-\sigma_2+2\tau_2) -\frac{1}{r}(2\rho_1-2\sigma_1-3\tau_1)-\frac{2}{r^2}(\rho-\sigma)=0, \qquad (3d)$$

$$\tau_{12}+\mu_{12}-\frac{1}{r}(\rho_2+\mu_2)-\text{ctg}\vartheta(\sigma_1-\tau_1)=0, \qquad (3e)$$

$$\sigma_{14}+\tau_{14}-\frac{1}{r}(2\rho_4-\sigma_4-\tau_4)=0, \qquad (3f)$$

$$\rho_{24}+\tau_{24}-\text{ctg}\vartheta(\sigma_4-\tau_4)=0. \qquad (3g)$$

Integrating (3f) and (3g) with respect to t and taking the functions of integration to vanish, we get

$$\sigma_1+\tau_1-\frac{1}{r}(2\rho-\sigma-\tau)=0, \qquad (4a)$$

$$\rho_2+\tau_2-\text{ctg}\vartheta(\sigma-\tau)=0. \qquad (4b)$$

Differentiating (4a), one obtains an equation which can be put into the form

$$\sigma_{11}+\tau_{11}-\frac{2}{r}(\rho_1-\sigma_1-\tau_1)=0, \qquad (5a)$$

while differentiating (4b) gives

$$\rho_{12}+\tau_{12}-\text{ctg}\vartheta(\sigma_1-\tau_1)=0. \qquad (5b)$$

Subtracting (5b) from (3e) and integrating with respect to ϑ (with the function of integration taken to vanish), one gets

$$\mu_1-\rho_1-\frac{1}{r}(\rho+\mu)=0, \qquad (4c)$$

and from this by differentiation one gets a relation which can be written

$$\mu_{11}-\rho_{11}-\frac{2}{r}\rho_1=0. \qquad (5c)$$

[1] H. Dingle, Proc. Natl. Acad. Sci. U. S. 19, 559 (1933).

Finally, by combining (3a), (3b), (3c), (3d), (5a), and (5c), one gets

$$\rho_{11}+\frac{2}{r}\rho_1+\frac{1}{r^2}\rho_{22}+\frac{\text{ctg}\vartheta}{r^2}\rho_2-\rho_{44}=0 \qquad (6)$$

that is, a wave equation for ρ.

The general solution of (6) is, of course, well known. It can be expressed, for example, as an expansion in spherical (zonal) harmonics. We limit ourselves here to the term of lowest order that can be expected to give a nontrivial result, that corresponding to an axial quadrupole source at the origin. The method can be extended to multipoles of higher order. We also take the solution to be sinusoidal in time and to represent an outgoing wave. Using the complex form (in the end the real part will be taken), the solution can be written

$$\rho=Ch(x)P_2(\cos\vartheta)e^{-i\omega t} \qquad (7)$$

where the frequency ω and C are constants, $x=\omega r$, and

$$h(x)=\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}[J_{\frac{5}{2}}(x)-iJ_{-\frac{5}{2}}(x)] \qquad (8)$$

with $J_n(x)$ denoting the Bessel function of order n. Once ρ has been found, the remaining unknowns can be determined by means of the field equations.

The constant C is to be determined in terms of the strength of the source. For this purpose we take a very simple model of the source, in the form of an infinitesimal system consisting of two massive particles situated on the Z axis and connected by a nearly mass-less spring. We assume that the energy-momentum-stress tensor has for its nonvanishing components in the Cartesian coordinate system T'_{33}, T'_{34}, and T'_{44}, with $x^3=z$. Representing the mass density of a particle by a Dirac delta function and neglecting the static part of the tensor, we take

$$T'_{44}=\tfrac{1}{2}p\frac{\partial^2\delta(\mathbf{r})}{\partial z^2}, \qquad (9)$$

where

$$p=p_0e^{-i\omega t}, \qquad (10)$$

and $\delta(\mathbf{r})$ is the three-dimensional Dirac delta function. From (9) it follows that

$$p=\int z^2T'_{44}d\tau. \qquad (11)$$

The remaining components of the tensor are determined by integrating the equations of motion

$$T'^{\mu\nu}{}_{;\nu}=0, \qquad (12)$$

which in first approximation have the form

$$T'_{33,3}-T'_{34,4}=0, \qquad (13a)$$

$$T'_{43,3}-T'_{44,4}=0. \qquad (13b)$$

One obtains

$$T'_{33}=-\tfrac{1}{2}\omega^2p\delta(\mathbf{r}), \qquad (14a)$$

$$T'_{34}=-\frac{i}{2}\omega p\frac{\partial(\mathbf{r})}{\partial z}. \qquad (14b)$$

Transforming the components to the polar coordinate system and also expressing the three-dimensional delta function $\delta(\mathbf{r})$ in terms of the one-dimensional delta function $\delta(r)$, one gets the following nonvanishing components:

$$T_{11}=-\frac{1}{4\pi}\omega^2p\frac{\cos^2\vartheta}{r^2}\delta(r), \qquad (15a)$$

$$T_{12}=\frac{1}{4\pi}\omega^2p\frac{\cos\vartheta\sin\vartheta}{r}\delta(r), \qquad (15b)$$

$$T_{14}=-\frac{3}{4\pi}i\omega p\frac{\cos^2\vartheta}{r^3}\delta(r), \qquad (15c)$$

$$T_{22}=\frac{1}{4\pi}\omega^2p\sin^2\vartheta\delta(r), \qquad (15d)$$

$$T_{24}=\frac{3}{4\pi}i\omega p\frac{\sin\vartheta\cos\vartheta}{r^2}\delta(r), \qquad (15e)$$

$$T_{44}=\frac{3}{4\pi}p\frac{5\cos^2\vartheta-1}{r^4}\delta(r). \qquad (15f)$$

To take account of the source, (1) should be replaced by

$$G_{\mu\nu}=-T_{\mu\nu}, \qquad (16)$$

in units such that $8\pi k/c^2=1$. If one follows through the steps previously used in obtaining the wave equation for ρ, (6), one finds that now the corresponding equation has the form

$$\nabla^2\rho+\omega^2\rho=\frac{3}{2\pi}p\left[\frac{5\cos^2\vartheta-4}{r^4}+f(r)\right]\delta(r), \qquad (17)$$

where $f(r)$ is an arbitrary function. On the other hand, the solution for ρ previously obtained, as given in (7), for small values of r can be written in the form

$$\rho\simeq-\frac{3iC}{2\omega^3}\frac{\partial^2\left(\frac{1}{r}\right)}{\partial z^2}e^{-i\omega t}, \qquad (18)$$

from which one gets

$$\nabla^2\rho\simeq\frac{9iC}{\omega^3}\frac{5\cos^2\vartheta-1}{r^4}\delta(r)e^{-i\omega t}. \qquad (19)$$

Comparing (17) with (19) and using the expression for p given by (10), one finds that

$$C=-\frac{i\omega^3}{6\pi}p_0. \tag{20}$$

The solution for all four unknowns, expressed in terms of real functions, is found to be given by

$$\rho=-\frac{\omega^3}{6\pi}p_0[j_2(x)\sin\omega t+j_{-2}(x)\cos\omega t]P_2(\cos\vartheta), \tag{21a}$$

$$\sigma=-\frac{\omega^3}{12\pi}p_0\{[(3k_2(x)+j_2(x))\sin\omega t+(3k_{-2}(x)+j_{-2}(x))\cos\omega t]P_2(\cos\vartheta)-(k_2(x)+j_2(x))\sin\omega t-(k_{-2}(x)+j_{-2}(x))\cos\omega t\}, \tag{21b}$$

$$\tau=-\frac{\omega^3}{12\pi}p_0\{[(k_2(x)-j_2(x))\sin\omega t+(k_{-2}(x)-j_{-2}(x))\cos\omega t]P_2(\cos\vartheta)+(k_2(x)+j_2(x))\sin\omega t+(k_{-2}(x)+j_{-2}(x))\cos\omega t\}, \tag{21c}$$

$$\mu=-\frac{\omega^3}{6\pi}p_0[(2q_2(x)+j_2(x))\sin\omega t+(2q_{-2}(x)+j_{-2}(x))\cos\omega t]P_2(\cos\vartheta). \tag{21d}$$

Here

$$j_n(x)=\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}J_{n+\frac{1}{2}}(x), \tag{22a}$$

$$k_n(x)=\frac{1}{x}\int_\infty^x j_n(u)du, \tag{22b}$$

$$q_n(x)=x\int_\infty^x u^{-2}j_n(u)du. \tag{22c}$$

In the limit of a static system ($\omega=0$) one finds that μ is proportional to the Newtonian potential, as is to be expected. It will be noted that in the above solution ρ, μ, and $\sigma+\tau$ are of a form corresponding to a quadrupole wave, i.e., they depend on the angle through $P_2(\cos\theta)$. On the other hand, the dependence of $\sigma-\tau$ on the angle is given by $\sin^2\theta$.

We see then that, in the linear approximation, the field equations admit a solution describing gravitational waves emitted by an oscillating physical system. A calculation shows that these waves are real, that is, the Riemann-Christoffel tensor is different from zero, so that they cannot be removed by a coordinate transformation.

To calculate the rate of emission of energy by the material system, one can make use of the gravitational energy-momentum-stress pseudotensor t_ν^μ. For this it is desirable to go over to a Cartesian coordinate system, since in that case the first-order terms in the metric tensor will give second-order terms in t_ν^μ, which will be the terms of lowest order, and these will not depend on terms of higher order in the metric tensor. The calculation is somewhat tedious, and the final result for the rate of energy emission is

$$N=\frac{1}{120\pi}\omega^6p_0^2,$$

in the present units, or

$$N=\frac{k}{15c^5}\omega^6p_0^2,$$

in the usual units. This result agrees with that obtained by others.[2]

It is planned to use the above solution as the starting point for a more accurate calculation. The interesting question is whether the exact equations have a solution going over into the above for sufficiently weak fields.

ACKNOWLEDGMENTS

In conclusion, we should like to express our indebtedness to Professor G. Racah for helpful advice and discussion.

[2] L. Landau and E. Lifshitz, *Classical Theory of Fields* (Addison-Wesley Press, Cambridge, 1951), p. 331.

Dynamics of a Lattice Universe by the Schwarzschild-Cell Method

RICHARD W. LINDQUIST* AND JOHN A. WHEELER

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

1. INTRODUCTION AND SUMMARY; THE SCHWARZSCHILD-CELL METHOD

MANY problems of electrostatics can be expressed in terms of the elementary $1/r$ potential due to a point charge. Gravitation theory provides the analog of the $1/r$ potential in Schwarzschild's expression

$$ds^2=\sum_{\alpha\beta} g_{\alpha\beta}dx^\alpha dx^\beta$$
$$=(1-2Gm/c^2r)^{-1}dr^2+r^2[d\theta^2+\sin^2\theta d\varphi^2]$$
$$-(1-2Gm/c^2r)dT^2 \quad (1)$$

for the 10 gravitational or metric potentials of Einstein for the effect of an elementary concentration of mass in an asymptotically flat space-time continuum. Here G is the Newtonian constant of gravitation, 6.67×10^{-8} cm^3/g sec^2, c the velocity of light (cm/sec), and m the mass in grams. The equations of gravitation theory are nonlinear. The principle of superposition, so central in electrostatic theory, does not apply. Consequently, the Schwarzschild solution has no such far-reaching realm of application as the Coulomb potential. Nevertheless there exists a class of problems that receive a simple approximate solution by way of the Schwarzschild formula.

We consider a number of mass concentrations so distributed in space and of such relative magnitudes, that the zone of influence of each can be reasonably approximated by a sphere. Inside each cell we replace the actual gravitational potentials by the expressions of Schwarzschild. This approximation demands that the distribution of gravitational influences just external to each sphere should depart relatively little from spherical symmetry. Such a treatment is inspired by the success of Wigner and Seitz[1] in analyzing electronic wave functions in crystal lattices. They approximate the elementary lattice cell by a sphere of the same volume. The wave equation for a problem of nonseparable coordinates becomes separable. Applied to a problem where the exact solution is known—where the potential is constant—the accuracy of this analysis has been tested with favorable results.[2]

From analogy with the Wigner-Seitz analysis one might expect that the gravitational potentials ought to have zero normal derivative at the boundary of each lattice cell, idealized as spherical. The Schwarzschild potentials do not satisfy this requirement. The derivative $(d/dr)(1-2Gm/c^2r)$ is proportional to the Newtonian gravitational field, Gm/r^2, and does not go to zero at a finite distance. This derivative measures the rate of acceleration of an infinitesimal test body placed at the point in question. This acceleration at the position of the cell boundary gives the acceleration of the cell boundary itself. Otherwise stated, the mass concentrations on either side of the cell boundary accelerate towards that boundary at such a rate as to nullify the discontinuity in matching of the normal derivative of the gravitational potentials that would otherwise occur. Thus the cell method in gravitation theory has an essentially new and dynamic feature that does not appear in the solid state problem. This new feature is simple and important. *It expresses the equation of motion of the mass at the center of a cell as a dynamic condition on the boundary of the cell.*

The rest of this paper elaborates this idea and applies it to the problem of the expanding universe. *The whole of the dynamics of the expansion and subsequent contraction is derived from the elementary static Schwarzschild solution* (*1*). For simplicity attention is limited to the case of equal masses arranged in a regular lattice in a closed space. Special interest attaches to the case of a very large number of masses, $N=600$, the largest number that can be arranged in a regular lattice. The calculated radius of maximum expansion, a_0, for this case agrees to 1.2% with the well-known result of Friedmann,

$$a_0=(4G/3\pi c^2)(\text{total mass}), \quad (2)$$

for the case of a homogeneous distribution of matter.

Briefly recapitulated, the Schwarzschild-cell method considers the dynamics of a lattice universe as a consequence of the field equations. These equations are fulfilled everywhere except at the interface between zones of influence, and fulfilled even there in an average way. Thus the field equations determine the equations of motion of the singularities that represent matter. In this sense we follow the spirit of Einstein, Infeld, and Hoffman.[3] However, our approximation method to

---

* National Science Foundation Predoctoral Fellow.

[1] E. P. Wigner and F. Seitz, Phys. Rev. 43, 804 (1933); *ibid.*, 46, 509 (1934); see also F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), Chap. 9.

[2] W. Shockley, Phys. Rev. 52, 286 (1937); F. C. Von der Lage and H. A. Bethe, Phys. Rev. 71, 612 (1947).

[3] Einstein, Infeld, and Hoffman, Ann. Math. 39, 66 (1939). See L. Infeld and A. Schild, Revs. Modern Phys. 21, 408 (1949) for the derivation of the equations of motion for an infinitesimal test particle moving at arbitrary speed; also L. Infeld and J. Plebanski, Bull. Acad. Polon. Sci., Class III, 4, 757 (1956).

determine the equations of motion of mass singularities from the field equations is quite different from theirs. They demand small relative velocities, and a space that is asymptotically flat, but make no symmetry requirements on the disposition of the masses. In contrast, symmetry alone is the demand of the method that decomposes all of curved space into Schwarzschild cells.

When closed space is divided up into infinitely many infinitely small cells one passes to the limit of the dust-filled universe of Friedmann. Section 2 summarizes the properties of this Friedmann solution and defines the ideas needed in comparing this uniform model with the lattice model. One introduces the idea of a "comparison uniformly curved space" or "comparison hypersphere." The boundaries of zones are marked on it that have the same reflection symmetries as do the zones of the lattice universe. Section 3 analyzes the geometry of these polyhedral zones on the comparison hypersphere. Section 4 derives the equation of motion of the boundary from two conditions: (1) the boundary of the Schwarzschild cell must be tangent to the comparison hypersphere; (2) the solid angle subtended by the Schwarzschild cell on the hypersphere must not change with time. To satisfy these two conditions, the radius of the Schwarzschild cell, and the radius of the comparison hypersphere must vary with time. The equation for the cell radius is found (Sec. 4) to be identical with that of a freely falling test particle. Section 5 shows that the radius of the comparison hypersphere has essentially the same time variation as does the radius of the ideal Friedmann universe. Section 6 analyzes possibilities for replacing the Schwarzschild cell approximation by an exact treatment of the lattice universe. Section 7 considers possible use of the variational principle for extensions of the method, and consequences of departures of a cell from perfect symmetry.

2. FRIEDMANN UNIVERSE AND THE LATTICE UNIVERSE COMPARED

A closed universe with zero cosmological constant in accord with the arguments of Einstein[4] is assumed. When this universe is homogeneous and isotropic, the metric can always be written[5]

$$ds^2=\frac{dr^2}{1-(r^2/a^2)}+r^2(d\theta^2+\sin^2\theta d\phi^2)-dT^2. \qquad (3)$$

The radius of curvature, $a(T)$, depends upon time in a way that is governed by the equation of state. In the Friedmann case, where the universe is filled with a uniform dust of total mass M and zero pressure, the radius satisfies the equation of energy,

$$Mc^2+Mc^2(da/dT)^2-(4GM^2/3\pi a)=0. \qquad (4)$$

[4] A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, New Jersey, 1950), p. 107.

[5] H. P. Robertson, Revs. Modern Phys. **5**, 62 (1933).

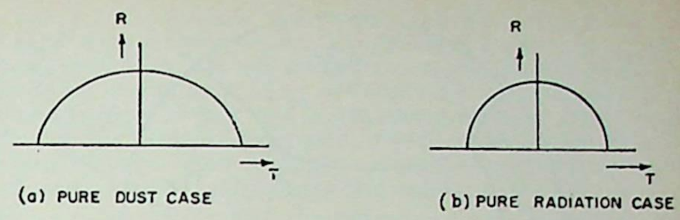


FIG. 1. Radius of the universe, a, as a function of cotime, $T=$(velocity of light)·(time), for a homogeneous isotropic universe. The system expands outward and falls back together again under gravitational attraction, but the detailed dynamics depends upon the equation of state of the medium. In Friedmann's universe (case a) the pressure is zero. In the case of pure radiation (case b) the pressure is equal to one third of the energy density, and the circle $a^2+T^2=a_0^2$ takes the place of Friedmann's cycloid. The intermediate case has been treated numerically by G. Gamow, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. 27, No. 10 (1953).

The maximum radius is

$$a_0=4GM/3\pi c^2. \qquad (2a)$$

The relation between radius, a, and cotime, T $=$(velocity of light)·(time), is a cycloid (Fig. 1) given parametrically in the form

$$a=(a_0/2)(1+\cos\eta)$$
$$T=(a_0/2)(\eta+\sin\eta). \qquad (5)$$

The three dimensional curved space of Friedmann at any one cotime, T, is equivalent to the three-dimensional hypersurface of a sphere in a four-dimensional *Euclidean* space. The extra dimension has no physical meaning; it is an imbedding dimension. Introducing Cartesian coordinates,

$$u_1=a\sin\chi\sin\theta\sin\varphi,$$
$$u_2=a\sin\chi\sin\theta\cos\varphi,$$
$$u_3=a\sin\chi\cos\theta,$$
$$u_4=a\cos\chi,$$

and letting

$$r=a\sin\chi, \qquad (6)$$

gives the metric (3) the simple form

$$ds^2=du_1^2+du_2^2+du_3^2+du_4^2-dT^2.$$

On this sphere we mark out the vertices of a regular figure to give a geometry that can be compared with the geometry of the lattice universe. Particular dust particles specify these N vertices. Every vertex can be equidistant from its nearest neighbors only when $N=5$, 8, 16, 24, 120 or 600.[6]

The case $N=8$ gives one of the simplest of these arrangements. For any typical one, P, of the eight points there is another one, A, of the eight which may be called its antipode. The remaining six particles are most conveniently named as N(north), S(south), E, W,

[6] See H. M. S. Coxeter, *Regular Polytopes* (Methuen and Company, Ltd., London, 1948), in particular Table I (ii) on pp. 292–293, for an enumeration of the regular polytopes in four dimensions.

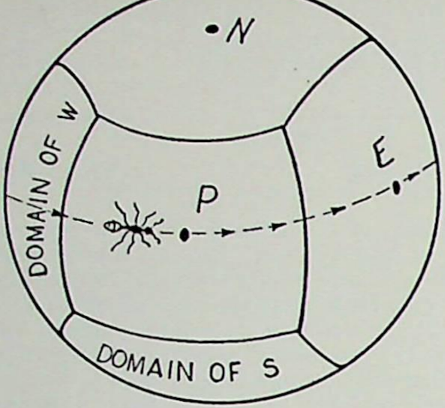


FIG. 2. Two-dimensional analog of a closed lattice universe. The third dimension in the drawing merely provides a dimensionality high enough to allow imbedding a two-dimensional space with the topology of a sphere. The superfluous radial coordinate has no relation to the time coordinate. There are six lattice cells: P, E, A (antipode), W and N, S. Each is like a square in its count of edges and corners but is a deformed square. The distortion is such that the angle at a corner is 120° instead of the normal 90°. In terms of distances:

All distances in terms of radius, a, of sphere	Center of cell to center of edge	Center of cell to corner
Lattice cell of Fig. 2	$\frac{2\pi a}{8}=0.7854a$	$a\sin^{-1}\frac{6^{\frac{1}{2}}}{3}=0.9553a$
Circular domain on surface of sphere with same center and same area	$a\cos^{-1}\left(\frac{2}{3}\right)=0.8411a$	
Ratios	0.9338	1.1359
Square of same area in flat space $\left(=\frac{4\pi}{6}a^2\right)$	$\left(\frac{\pi}{6}\right)^{\frac{1}{2}}a=0.7236R$	$\left(\frac{\pi}{3}\right)^{\frac{1}{2}}a=1.0233a$
Circle of same area in flat space	$\left(\frac{2}{3}\right)^{\frac{1}{2}}a=0.8165a$	
Ratios	0.8862	1.2533

U(up) and D(down). Proceeding from P to any one of these six nearest neighbors, say E and continuing on in the same direction, then one comes next to the antipodal mass, A; then to W and finally back to P. Likewise, proceeding upward from P, one comes in turn to U, A, D and back to P; and so on. Figure 2 sketches the analog of this lattice arrangement for a two-dimensional closed surface imbedded in a three-dimensional Euclidean space.

An infinitesimal test particle ordinarily lies closer to one of the lattice centers, P, than to another. It may be said to belong to the zone of influence of P, or to the *lattice cell* centered on P. Between P and E lies a two-dimensional array of points which are equidistant from P and E: the *boundary* or *interface* between the domains of P and E.

From Friedmann's simple metric with its uniform curvature we turn to the metric of a lattice universe, where all the mass is concentrated into N centers. The curvature of the metric, R_{ijkl}, now varies from place to place as indicated qualitatively in Fig. 3. However, the contracted curvature tensor, $R_{ik}=g^{\alpha\beta}R_{i\alpha k\beta}$ vanishes throughout the mass-free space between the vertices. Moreover the symmetry group of the vertices is completely unchanged, as are the reflection symmetries of the metric at the cell boundaries.

How is one to deal with this metric with its complicated variation in space? Can one simplify the problem by limiting attention to those parts of space that lie infinitesimally close to zonal boundaries, with their reflection symmetries? No, even the metric in these limited regions of space does not conform to the surface of a comparison hypersphere. One naturally chooses the radius of this comparison hypersphere so that points halfway between nearest neighbors will lie outside the comparison hypersphere, and points equidistant from three nearest neighbors will lie inside.

To simplify the metric in an individual cell we have to simplify the geometry of its boundary. We *replace the typical cell of the lattice universe by a Schwarzschild cell* that (1) has a spherically symmetric metric:

$$ds^2=[1-(2m^*/r)]^{-1}dr^2+r^2[d\theta^2+\sin^2\theta(d\varphi)^2] -[1-(2m^*/r)]dT^2, \quad (7)$$

$m^*=$ mass of one singularity expressed in units of length,

$$=(G/c^2)(\text{mass of singularity})=Gm/c^2, \text{ and } m=(\text{total mass})/(\text{number of vertices})=M/N; \quad (8)$$

(2) has a spherically symmetric boundary whose radius r depends on time, and (3) conformally joins at this radius r onto a *comparison hypersphere* of a radius, a, that also depends on time. This radius defines what we mean by the radius of the lattice universe. To state in quantitative terms this joining or *tangency* condition now requires examination of geometry of the comparison hypersphere.

There is a second motivation to study geometry on the hypersphere. There one can define polyhedral zones and compare them with spherical zones to test quantitatively how nearly the dimensions of the two objects agree. Such a comparison is not possible for the lattice space itself because its metric is not known. Therefore the analogous comparison on the surface of the hypersphere—and in flat space—supply our only simple means to estimate the accuracy of the Wigner-Seitz approximation.

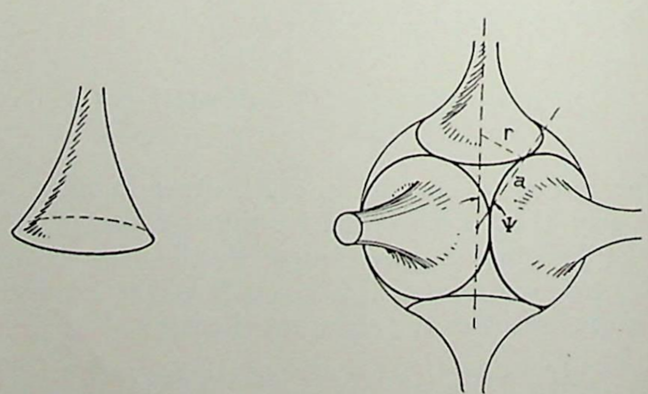


FIG. 3. Qualitative character of the space in a lattice universe. The space is closed up, but not by an everywhere uniform curvature as in the Friedmann universe.

3. GEOMETRY OF SPHERICAL AND POLYHEDRAL CELLS IN UNIFORMLY CURVED SPACE

The shape of the typical lattice cell in uniformly curved space is easily visualized by considering the two-dimensional analog (Fig. 2) of the three-dimensional lattice universe containing eight particles. From that analogy, or from direct analysis, one sees that the faces of the cube-like lattice cell meet at a dihedral angle of 120°, rather than the familiar 90° (Figs. 2 and 4). An edge is common to three cells, not to the four expected from the familiar picture of stacked blocks. At a corner only four lattice cells meet, instead of the usual eight; and only four edges, instead of the usual six. These four edges radiate from the corner at the regular tetrahedral angles. The corner of a lattice cell appears blunt when compared with the corner of a cube. To approximate the lattice cell as a sphere is expected to be slightly better than the replacement of a cube by a sphere as done in solid state physics. This conclusion is borne out by the comparison of distances given in Table I.

Similar distance comparisons for the other regular lattices in a space of uniform curvature appear in Table II. Thus the lattice cell of a spherical polytope

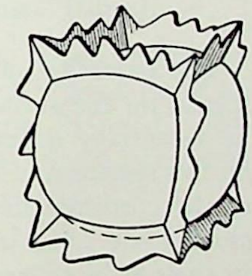


FIG. 4. Shape of typical cell in case of eight particle lattice universe. The cell is like a deformed cube. Three cells meet at an edge rather than the four of Euclidean geometry, and at a corner four cells meet rather than six.

deviates from a sphere considerably less than does its counterpart in flat space. This makes it reasonable to develop the spherical cell approximation.

To obtain a measure of the size of a spherical cell, both absolute and relative to the total hypersurface, we imbed the curved three-dimensional space at a fixed moment of time into a Euclidean four-dimensional space. Denote by a the radius of the hypersphere and by r_0 the radius of the spherical cell, both as measured in the flat imbedding space. Denote by ψ the angular separation of the center of the spherical cell and its boundary as seen from the center of the hypersphere:

$$\sin\psi = r_0/a. \tag{9}$$

The cell subtends at the center of the hypersphere a hypersolid angle

$$\Psi = \frac{\int d(\text{three-dimensional volume})}{a^3} = \int_{\text{inside}} \sin^2\chi \sin\theta d\chi d\theta d\varphi = \pi(2\psi - \sin 2\psi) \tag{10}$$

TABLE I. Comparison of (1) lattice cell dimensions for the case of eight mass centers with (2) dimensions of a spherical cell of the same volume in curved space, under assumption of *uniform* curvature of the space, i.e., uniform mass density. In the case when the mass of each cell is concentrated at the center of that cell, the geometry of space is changed—and changed in a time dependent way. Then these dimensions no longer apply exactly, and have primarily only illustrative value.

All distances in terms of radius of curvature, a, of space	Center of cell to center of face	Center of cell to center of edge	Center of cell to corner
Lattice cell of volume $(1/8)2\pi^2a^3$	$0.785a$	$0.955a$	$1.047a$
Spherical domain in curved space with same center and volume		$0.883a$	
Ratios	0.889	1.082	1.186
Cube of same volume in flat space	$0.676a$	$0.955a$	$1.170a$
Sphere of same volume in flat space		$0.838a$	
Ratios	0.806	1.140	1.396

in analogy to the familiar formula in three-dimensional geometry

$$\Omega = 2\pi(1-\cos\psi)$$

for solid angle subtended by a cone of half-angle ψ. Out of the whole hypersurface or three volume of the hypersphere,

$$2\pi^2a^3, \tag{11}$$

the spherical cell occupies the fraction

$$\Psi/\Psi_{\text{total}} = (2\psi - \sin 2\psi)/2\pi. \tag{12}$$

We wish to cut away the uniformly curved space occupied by a spherical cell and replace that cell by a Schwarzschild cell. In the Schwarzschild cell the metric is nonuniform but still spherically symmetrical. We want the Schwarzschild cell to have the same spherical boundary as the uniformly curved spherical cell that it replaces, and to be "tangent" to the hypersphere at the point of join.

How big shall the spherical boundary be chosen? Two alternative conditions suggest themselves for defining the angle ψ:

TABLE II. Distance ratios as measures of the "roundness" of a lattice cell in curved space. For comparison the same distance ratios are also given for flat space for a cell with the same number of faces, edges and corners. The cells are most nearly spherical when there are 120 mass centers.

Number of identical cells	Name that would be given to one such cell in flat space	Ratio distance to: face or corner / "radius" of sphere of same volume			
		Curved space	Flat space	Curved space	Flat space
5	tetrahedron	0.862	0.671	1.246	2.013
8	cube	0.889	0.806	1.186	1.396
16	tetrahedron	0.763	0.671	1.525	2.013
24	octahedron	0.880	0.846	1.320	1.465
120	dodecahedron	0.917	0.910	1.133	1.146
600	tetrahedron	0.679	0.671	1.947	2.013

Condition I: The boundary sphere shall cut out of the hypersphere $1/N$ of the total solid angle.

Condition II: The boundary sphere associated with one mass concentration shall just touch the boundary spheres of all its nearest neighbors.

Condition I is the direct generalization to curved space of the standard Wigner-Seitz approximation. The boundary angle ψ is fixed by

$$\frac{\Psi}{\Psi_{\text{total}}}=\frac{1}{N}=\frac{2\psi-\sin 2\psi}{2\pi}. \tag{13}$$

N spherical cells, each of which will fit into $1/N$th the hypersurface—or three volume—of the hypersphere will in some places overlap and in other places leave regions of "no man's land." When the cells are merely cutouts from the hypersurface, they have uniform curvature and in the regions of overlap meet every reasonable requirement of "tangency." Consider the other case when the regions inside the boundary spheres are Schwarzschild cells, *adjusted here and always to tangency to the comparison hypersphere at their spherical boundaries*. Where the boundaries of two spherical cells interact, the two Schwarzschild metrics are tangent because at that point both are tangent to the hypersphere. Where the cells overlap, the intersecting Schwarzschild metrics depart from tangency in one sense; and they depart from tangency in the opposite sense wherever the intersection lies outside the spherical boundaries of the cells. Condition I makes the metrics of two Schwarzschild cells be tangent to each other in an average way, it being understood that at the spherical boundary of each cell the Schwarzschild metric has been made tangent to the comparison hypersphere in a way defined later.

In contrast, Condition II gives a matching only at the one point of contact midway between neighboring mass centers. Away from this point of contact, in the now enlarged "no man's land" between spherical cells, two extrapolated Schwarzschild metrics at their intersection depart more and more from tangency. The departure always has the same sign. Consequently, Condition II will be a less reasonable criterion for cell size than Condition I. Thus far, we have examined the two criteria for cell size from the point of view of matching of boundary conditions thinking of the metric inside as of Schwarzschild form. Now imagine instead a uniformly curved space inside of these boundaries. Then the tangency requirements are automatically satisfied. Another point of view offers itself to compare the two criteria for cell size. Table III shows that the polyhedral cell dimensions are not as well matched by Condition II as those spheres of Condition I. One is not in a position so easily to calculate distances when the metric has the nonuniform curvature of the true lattice space, but it is reasonable to believe that Condition I continues to give the more reasonable cell dimensions, as it also better satisfies on the average the tangency requirement.

We have now decided that the spherical boundary of one Schwarzschild cell is to cut out $1/N$th of the volume of the comparison hyperspace (13). Next we formulate the requirement that the cell be *tangent* to the hypersphere at the point of join. Imagine a surveyor told to find the fraction of the earth's area enclosed within a circular kingdom, but only allowed to make measurements on or near its periphery. His solution is simple: Measure the circumference of the kingdom, and compare it with that of an infinitesimally smaller circle also on the earth's surface. Also measure the difference in their radii. If the earth were flat, one would find

$$\frac{d(\text{circumference})}{d(\text{radial distance})}=2\pi.$$

However, due to its finite radius a, one finds

$$\frac{d(\text{circumference})}{d(\text{radial distance})}=\frac{d(2\pi a\sin\psi)}{a d\psi}=2\pi\cos\psi. \tag{14}$$

The extension of this method to the problem of a sphere on the surface of a given hypersphere requires no essentially different ideas. Choose any great circle on this sphere, and compare its circumference with that of the corresponding great circle on an infinitesimally smaller one. Let d (radial distance) denote the difference in their radii *as measured in an invariant manner along the surface of the hypersphere*. One is led again to (14). In view of the complete symmetry of the problem, one clearly obtains the same value of ψ regardless of which great circle is chosen for the measurement.

Thus there is an invariant way to say that one Schwarzschild cell shall cut out $1/N$th of the tangent hypersphere, and to say this entirely in terms of *measurements in and near the cell boundary*:

$$\frac{1}{2\pi}\frac{d(\text{circumference})}{d(\text{radial distance})}=\cos\psi, \tag{15}$$

TABLE III. Comparison of polyhedral and spherical cells in the closed three space that forms the hypersurface of a sphere in 4 dimensional *Euclidean* space.

	Distance from center of cell in units of hypersphere radius				
				Radius of sphere that replaces cell according to	
N	Center of face	Center of edge	Center of cell	Condition I	Condition II
5	0.912	1.150	1.315	1.057	0.912
8	0.785	0.955	1.047	0.883	0.785
16	0.524	0.785	1.047	0.686	0.524
24	0.524	0.617	0.785	0.595	0.524
120	0.314	0.365	0.388	0.343	0.314
600	0.135	0.232	0.388	0.199	0.135

where ψ is *defined* by

$$\frac{2\psi-\sin 2\psi}{2\pi}=\frac{1}{N}. \quad (16)$$

In addition, there is an invariant way to define the radius, a, of the hypersphere—or what we call the radius of the universe—in terms of measurements at the cell boundary:

$$a=\frac{(\text{cell radius})}{\sin\psi}=\frac{(\text{circumference of great circle})}{2\pi\sin\psi}. \quad (17)$$

4. DYNAMICS OF THE LATTICE DEFINED BY BOUNDARY CONDITIONS

The problem of matching metrics at the boundary between two cells has been replaced by the problem of matching the metric in one Schwarzschild cell to the metric of the comparison hypersphere. From this matching condition we now derive the change with time in the radius of the cell and the radius of the universe—in other words, the dynamics of expansion and recontraction—using (15), (16), and (17), which depend only on behavior of the metric near the cell boundary.

In the Schwarzschild metric (7), a great circle of radius r has the circumference $2\pi r$. A great circle of radius $r-\epsilon$ has the circumference $2\pi(r-\epsilon)$. The invariant infinitesimal distance between one circle and the other is

$$\epsilon[1-(2m^*/r)]^{-\frac{1}{2}}.$$

These two observations made near the boundary fix the fraction of the whole hypersurface occupied by one Schwarzschild zone:

$$\frac{1}{2\pi}\frac{d(\text{circumference})}{d(\text{radial distance})}=[1-(2m^*/r)]^{\frac{1}{2}}=\cos\psi_N, \quad (18)$$

where ψ_N is fixed by N, according to (16). Thus, the radius r of the Schwarzschild cell cannot depend upon time, and we end up with a universe that has no dynamics at all! Obviously we have made an oversight somewhere.

The difficulty arises because the two circles were compared at equal values of the Schwarzschild time coordinate. We tacitly assumed that the Schwarzschild coordinates, r and T, are well adapted to describing the space-time continuum inside one cell, but they are not. Their inappropriateness appears even more clearly for r than for T. As the size of the lattice cell grows or shrinks, the range of the r coordinate increases or decreases. It appears as if one has to add on or chop off space at the cell boundary! It is much more natural to think of a new pair of space and cotime coordinates, ρ and τ, which are functions of r and T with the following properties. The boundary of the lattice cell is described by the statement that $\rho=\rho(r,T)=$constant at that point. The point $\rho-\epsilon$ in the "coordinate patch" of one lattice cell is to be identified with the point $\rho+\epsilon$ in the extrapolated coordinate patch of another lattice cell. This identification is not to be made for equal values of the Schwarzschild time T because then the two coordinate patches do not *overlap* but only *intersect* (Fig. 5). This failure to overlap can be stated this way: The coordinate surface, $T=$constant, does not stand perpendicularly to the boundary surface, $\rho=$constant, but *obliquely* to it. The appropriate new time variable, $\tau=\tau(r,T)$, has the property that the surface, $\tau=$constant, *does* stand perpendicularly to the boundary surface, $\rho=$constant. This orthogonality of the two surfaces is demanded by the condition of mirror symmetry at the zone boundary.

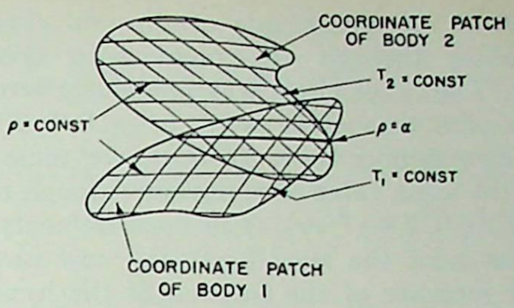


FIG. 5. The coordinate patches of two lattice cells merely intersect at the boundary, $\rho=\alpha$, and do not overlap as they should when the surfaces in question belong—as in the figure—to equal values of the Schwarzschild time, $T_1=T_2=$const. Real overlap or tangency demands instead equal values of a new time coordinate τ, not shown. Surfaces of constant τ intersect the zone boundary perpendicularly; surfaces of constant T, like those shown here, make oblique intersections.

We later construct a coordinate system, ρ, τ, that satisfies the orthogonality requirement. However, for formulation of the boundary condition at the surface such detail is not needed. It is enough to demand that the two circles of slightly different sizes be compared at the same value of the τ coordinate; in other words, that the ring-like surface connecting the two circles shall be orthogonal—in the four-dimensional sense—to the boundary of the lattice cell. This condition, being invariant, can also be formulated in the original Schwarzschild coordinate system.

The boundary will move in the interval of *Schwarzschild* cotime dT from r to $r+dr$. Describe this change by the four vector $(dr,0,0,dT)$ (in contravariant components). At this stage the ratio, dr/dT, of the components of this vector is not known. Construct a new vector that is (1) perpendicular to this vector, and (2) perpendicular to the two independent vectors that lie in the boundary $(0,d\theta,0,0)$ and $(0,0,d\varphi,0)$. The new vector is uniquely determined up to a multiplicative factor, and in one convenient normalization has the contravariant components

$$(-g^{rr}dT,0,0,g^{TT}dr).$$

This vector, being normal to the boundary, lies in the surface of constant τ.

Return to the comparison of the two circles. One circle passes through the point with coordinates $(r,0,0,T)$. The other circle was previously erroneously taken to pass through a point $(r-\epsilon, 0, 0, T)$ at the *same* Schwarzschild time. Instead, we must take a circle at the same value of τ, passing through the point $(r-g^{rr}dT, 0, 0, T+g^{TT}dr)$. The measurements on the two circles near the zone boundary now provide an invariant measure of the fraction of the hypersphere spanned by the Schwarzschild cell:

$$\cos\psi_N = \frac{1}{2\pi}\frac{d(\text{circumference})}{d(\text{radial distance})}$$

$$= \frac{g^{rr}dT}{[g_{rr}(g^{rr}dT)^2 + g_{TT}(g^{TT}dr)^2]^{\frac{1}{2}}}$$

$$= \frac{g^{rr}dT}{[g^{rr}g^{TT}]^{\frac{1}{2}}[g_{TT}(dT)^2 + g_{rr}(dr)^2]^{\frac{1}{2}}}$$

$$= \left(\frac{g^{rr}}{g^{TT}}\right)^{\frac{1}{2}} \frac{dT}{d(\text{proper cotime})}. \tag{19}$$

The angle ψ_N and the hypersolid angle Ψ spanned by the cell must remain constant in time; *the motion of the boundary must adjust itself to this match-up requirement.* Therefore (19) determines the equation of motion of the boundary:

$$\frac{dT}{d(\text{proper cotime})} = \frac{\cos\psi_N}{[1-(2m^*/r)]} \tag{20}$$

or

$$(dr/dT) = \pm(\cos\psi)^{-1}[1-(2m^*/r)][(2m^*/r) - \sin^2\psi]^{\frac{1}{2}}. \tag{21}$$

This completes the derivation of the dynamics of expansion from the metric match-up conditions.

The equation of motion of the boundary, (20) or (21), is identical with the law of conservation of energy for a unit test particle thrown out radially from the mass, m. The expression on the left side of (20)—after multiplication by c^2—gives the rest plus kinetic energy of the test particle,

$$\frac{c^2}{[1-(v^2/c^2)]^{\frac{1}{2}}} = c^2\frac{dT}{d(\text{proper cotime})}, \tag{22}$$

and the expression on the right similarly gives the relativistic generalization of its total energy diminished by its potential energy in the gravitational field. This integrated form of the equation of motion is less familiar than the differential form, analogous to Newton's equation of motion,

$$d^2x^i/ds^2 + \Gamma_{\alpha\beta}{}^i(dx^\alpha/ds)(dx^\beta/ds) = 0 \tag{23}$$

or

$$d^2T/ds^2 + g^{TT}(dg_{TT}/dr)(dr/ds)(dT/ds) = 0 \tag{24}$$

[compare (20)], and

$$d^2r/ds^2 + \tfrac{1}{2}g^{rr}(dg_{rr}/dr)(dr/ds)^2 - \tfrac{1}{2}g^{rr}(dg_{TT}/dr)(dT/ds)^2 = 0, \tag{25}$$

[compare (21)]. However the content is the same: the entire time dependence of the boundary between two lattice cells can be described by the behavior of a particle that falls towards both mass points simultaneously under the action of their gravitational attractions.

The integrated form of the equation of motion, (21), is less complete than the differential form because it possesses the singular solution

$$r = 2m^*/\sin^2\psi_N = \text{a constant}. \tag{26}$$

This singular solution already appeared (18) in the first, improperly formulated, attempt to derive the dynamics of the lattice cell from conditions at the boundary. Similar examples of singular solutions are common in classical physics, if one deals only with first integrals of the equations of motion. For example, a particle tossed straight up in a uniform gravitational field, that came to rest at the top of its flight—*and stayed at rest*—would satisfy the law of conservation of energy. However, it would violate Newton's second law of motion. Likewise an infinitesimal test particle permanently at the location (26) cannot satisfy the geodesic equation of motion (23). In mathematical terms, (26) is a singular solution of equations in the sense that it represents the *envelope* of the trajectories of test particles which reach the same distance of maximum excursion r at different values of T.

5. RADIUS OF THE LATTICE CELL AND OF THE UNIVERSE AS FUNCTIONS OF THE TIME

Equation (21) describes a motion of the cell boundary that changes from expansion to contraction when the last term vanishes; that is, when the cell radius, r, reaches a maximum value, R, defined by

$$r_{\max} \equiv R = 2m^*/\sin^2\psi. \tag{27}$$

TABLE IV. Maximum radius, a_0, attained during expansion of the lattice universe (as defined and calculated by the Schwarzschild-cell approximation method) compared with Friedmann's value for the maximum radius of a universe uniformly filled with dust. The radius is expressed relative to the Schwarzschild "radius" $2M^* = 2GM/c^2$ *defined* by the mass of the whole universe and also relative to the Schwarzschild radius $2m^* = 2Gm/c^2 = 2M^*/N$ associated with the mass of one lattice cell.

		Condition I			Condition II	
N	ψ	$a_0/2m^*$	$a_0/2M^*$	ψ	$a_0/2m^*$	$a_0/2M^*$
5	60.59°	1.5126	0.3025	52.24°	2.0239	0.4048
8	50.60°	2.1672	0.2709	45°	2.9293	0.3535
16	39.34°	3.9256	0.2454	30°	8.0000	0.5000
24	34.10°	5.6750	0.2365	30°	8.0000	0.3333
120	19.63°	26.376	0.2198	18°	33.889	0.2824
600	11.42°	128.82	0.2147	7.66°	406.06	0.6768
∞[a]	0°		0.2122	0°		0.2122

[a] Friedmann universe.

In terms of $r_{\max}$ the equation of motion of the boundary takes the form

$$dT=\frac{\pm[1-(2m^*/R)]^{\frac{1}{2}}dr}{[1-(2m^*/r)][(2m^*/r)-(2m^*/R)]^{\frac{1}{2}}}. \quad (28)$$

Direct integration leads to a closed formula[7] connecting the cell radius with the Schwarzschild cotime T:

$$T=\left[\frac{(R-2m^*)(R-r)r}{2m^*}\right]^{\frac{1}{2}} +\left[\frac{R-2m^*}{2m^*}\right]^{\frac{1}{2}}(R+4m^*)\text{ arc cos}\left(\frac{r}{R}\right)^{\frac{1}{2}} +2m^*\ln\frac{[r(R-2m^*)]^{\frac{1}{2}}+[2m^*(R-r)]^{\frac{1}{2}}}{[r(R-2m^*)]^{\frac{1}{2}}-[2m^*(R-r)]^{\frac{1}{2}}}. \quad (29)$$

More interesting is the connection between the cell radius and the proper cotime τ measured on an infinitesimal test particle that sits at the zone interface, falling simultaneously towards both mass centers:

$$d\tau=\frac{dT}{(dT/d\tau)} =\frac{dr}{[(2m^*/r)-(2m^*/R)]^{\frac{1}{2}}}, \quad (30)$$

a connection which integrates to

$$\tau=(R/2m^*)^{\frac{1}{2}}[r^{\frac{1}{2}}(R-r)^{\frac{1}{2}}+R\text{ arc cos}(r/R)^{\frac{1}{2}}]. \quad (31)$$

This is the equation for a cycloid, as appears from the parametric representation

$$\left.\begin{aligned} r&=\tfrac{1}{2}R(1+\cos\eta)=(M^*/N\sin^2\psi_N)(1+\cos\eta)\\ \tau&=\tfrac{1}{2}R(R/2m^*)^{\frac{1}{2}}(\eta+\sin\eta)\\ &=(M^*/N\sin^3\psi_N)(\eta+\sin\eta)\end{aligned}\right\}. \quad (32)$$

The radius of the universe,

$$\begin{aligned} a=&\text{(radius of cell boundary in the imbedding Euclidean space)}/\sin\psi_N\\ =&(1/2\pi)\text{(circumference of great circle about this sphere)}/\sin\psi_N\\ =&r/\sin\psi_N \end{aligned} \quad (33)$$

is also obtained as a function of proper cotime in parametric form directly from (32),

$$a=(M^*/N\sin^3\psi_N)(1+\cos\eta). \quad (34)$$

The Schwarzschild-cell method predicts a cycloidal relation between radius of the universe and proper cotime which is identical to that found by Friedmann (5), except for the connection between maximum radius and mass. When the mass is distributed uniformly, as in Friedmann's model, the maximum radius has the value

$$a_0=4GM/3\pi c^2=4M^*/3\pi. \quad (35)$$

When the mass is concentrated into N centers, and when the concept of "radius of the universe" is identified with the concept of "radius of the hypersphere tangent to a Schwarzschild cell at a zone boundary" then the maximum radius has the value (34)

$$a_0=2M^*/N\sin^3\psi_N. \quad (36)$$

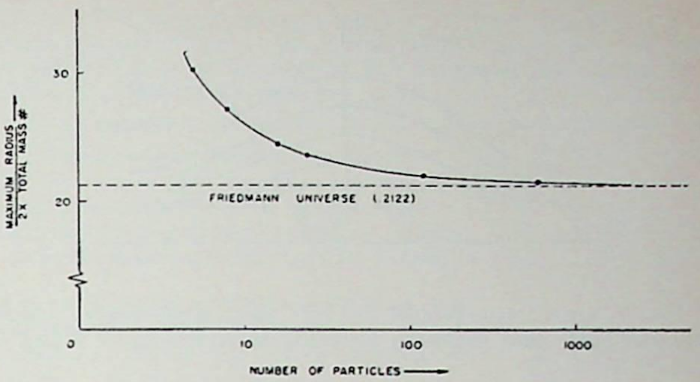


FIG. 6. Maximum radius of N-particle lattice universe, defined and derived in the Schwarzschild-cell approximation, compared with maximum radius of a universe filled with the same uniformly distributed mass.

Table IV lists the maximum radius calculated from (36) for the six regular lattice universes, using the two prescriptions discussed in Sec. 3 to define the size of the sphere that bounds a typical lattice zone. Condition I predicts a maximum radius that steadily approaches the Friedmann value with increasing N (Fig. 6) and that is only 1.2% above the Friedmann value for $N=600$. Thus, the Friedmann universe is the natural idealization of a universe containing a very large number of very small masses.

These results do not depend upon any particular choice of coordinate system. However, there is a coordinate system (ρ,θ,ϕ,τ) especially suited to describe dynamics within a Schwarzschild cell. Consider not merely one test particle that reaches its maximum r value, $r=R$, at $T=\tau=0$, but a whole collection of radially moving test particles that reach different maximum r values, R_1, R_2, $\cdots$, at $T=\tau=0$. Require that the geodesic of a test particle which is momentarily at rest at any point $r=R_i$ in a Schwarzschild field at time $t=\tau=0$ shall be given by $d\rho_i=0$. Thus ρ_i is to be interpreted as a "labeling coordinate" which distinguishes between the several test particles, and which depends on the value, R_i, of their maximum radii but not on their subsequent dynamics. For simplicity we choose the undetermined function $\rho_1(R_i)$ to be R_i itself. Also define the new cotime coordinate, τ, to be proper cotime measured along the appropriate test particle geodesic from the moment of maximum expansion. Then (29) and (31) define the transformation of coordinates $(r,T)\rightarrow(\rho,\tau)$ on

[7] Mr. Nicola N. Khuri kindly independently calculated this integral for us, which presumably also appears somewhere in the extensive literature of general relativity.

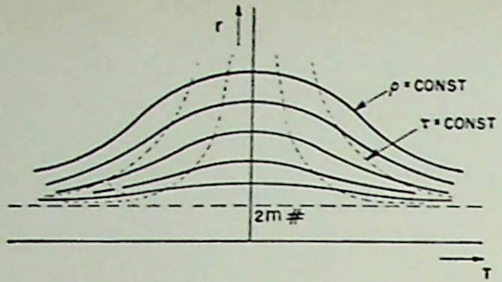


FIG. 7. Qualitative plot of the test particle geodesics $\rho=\text{const}$ and the space-like geodesics $\tau=\text{const}$ as functions of r and T.

replacing R by ρ. In the dimensionless units

$$r'=r/2m^* \quad \rho'=\rho/2m^*$$
$$T'=T/2m^* \quad \tau'=\tau/2m^*,$$

the transformation equations take the implicit form

$$T'=[(\rho'-1)(\rho'-r')r']^{\frac{1}{2}} +(\rho'-1)^{\frac{1}{2}}(\rho'+2)\text{ arc }\cos(r'/\rho')^{\frac{1}{2}} +\ln\frac{(r')^{\frac{1}{2}}(\rho'-1)^{\frac{1}{2}}+(\rho'-r')^{\frac{1}{2}}}{(r')^{\frac{1}{2}}(\rho'-1)^{\frac{1}{2}}-(\rho'-r')^{\frac{1}{2}}} \quad (37)$$

and

$$\tau'=(\rho')^{\frac{1}{2}}[(r')^{\frac{1}{2}}(\rho'-r')^{\frac{1}{2}}+\rho'\text{ arc }\cos(r'/\rho')^{\frac{1}{2}}]. \quad (38)$$

In the new dimensionless coordinates, ρ' and τ', the dimensionless metric, $ds'=ds/2m^*$, takes the form

$$(ds')^2=\left[\frac{(\rho'-r')}{4r'\rho'(\rho'-1)}\right]\left[(3\rho'-r')\left(\frac{r'}{\rho'-r'}\right)^{\frac{1}{2}} +3\rho'\text{ arc }\cos\left(\frac{r'}{\rho'}\right)^{\frac{1}{2}}\right]^2(d\rho')^2 +(r')^2[d\theta^2+\sin^2\theta d\varphi^2]-(d\tau')^2, \quad (39)$$

where r' is to be expressed in terms of ρ' and τ' from (38). The literature contains other coordinate systems, adapted to the study of other features of a Schwarzschild singularity.[8] The present coordinate system is well adapted to description of the expansion and recontraction process (Fig. 7).

6. RIGOROUS FORMULATION OF THE INITIAL VALUE PROBLEM

In principle, the problem of the dynamics of a closed N-body universe could be solved rigorously by first solving the Lichnerowicz[9] initial value equations, and then using the field equations to determine the metric at later times. This would provide a solution to the dynamical problem at least for a finite length of time.[10]

A class of solutions of the initial value problem has recently been found by Faurès and Misner,[11] under the assumption that $\partial g_{\mu\nu}/\partial T=0$, that is, that all mass centers are momentarily at rest on the initial surface $T=T_0$. These solutions provide an accurate description of the N-particle universe at the moment of maximum expansion. In isotropic coordinates, Misner's solution in the absence of charges is given by the expression

$$ds^2=\left(1+\sum_{i=1}^{N-1}\frac{\alpha_i}{|\mathbf{r}_i|}\right)^4(dx^2+dy^2+dz^2), \quad (40)$$

where $|\mathbf{r}_i|$ is the distance from the ith particle and $\alpha_i>0$. The Nth particle is located at infinity; in effect one is looking in through its Schwarzschild singularity at the rest of the universe. By comparing this expression with the standard Schwarzschild form in isotropic coordinates,

$$ds^2=\left(1+\frac{m^*}{2r}\right)^4(dx^2+dy^2+dz^2) -\left(1-\frac{m^*}{2r}\right)^2\left(1+\frac{m^*}{2r}\right)^{-2}dT^2, \quad (41)$$

as the distance $|\mathbf{r}_i|$ shrinks to zero, one finds the mass of the ith particle is given by

$$m_i^*=2\alpha_i\sum_{j\neq i}\left(1+\frac{\alpha_j}{|\mathbf{r}_{ij}|}\right) \quad \text{for} \quad i,j=1,2,\cdots,N-1$$
$$m_N^*=2\sum_i\alpha_i. \quad (42)$$

The location of the Schwarzschild singularity surrounding each particle can be defined as the surface of minimum area.

We have analyzed this metric in detail only for the five- and eight-body universes. To illustrate we deal only with the five-particle case. Take four masses at the points

$$(1,1,1),\ (1,-1,-1),\ (-1,1,-1)$$

and

$$(-1,-1,1).$$

The requirement of symmetry, and the condition that all five masses have the same value, gives the unique result

$$\alpha_i=2\sqrt{2};\quad m_i^*=16\sqrt{2} \quad \text{for all } i.$$

Hence a possible choice of metric for the moment of maximum expansion is

$$ds^2=\left(1+2\sqrt{2}\sum_{i=1}^{4}\frac{1}{|\mathbf{r}_i|}\right)^4(dx^2+dy^2+dz^2), \quad (43)$$

which rigorously satisfies the initial value requirements.

One can use this metric to estimate the validity of

[8] G. Lemaître, Ann. soc. sci. Bruxelles 53A, 51 (1933); J. L Synge, Proc. Roy. Irish Acad. A53, 83 (1950); A. Einstein and N. Rosen, Phys. Rev. 48, 73 (1935).

[9] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'electromagnétisme* (Masson et Cie, Paris, 1955).

[10] A. Raychaudhuri, Phys. Rev. 98, 1123 (1955); A. Komar, Phys. Rev. 104, 544 (1956).

[11] Y. Faurès-Bruhat, J. Rational Mech. Anal. 5, 951 (1956); C. W. Misner (to be published).

the two approximate match-up conditions (Conditions I and II) of Sec. 3. The most convenient quantity to compute for comparison with the rigorous solution is the invariant distance $2d$ between two Schwarzschild singularities, as measured along a geodesic between them. For the case of a pure Schwarzschild field this distance is given by

$$\frac{d}{2m^*}=\frac{1}{2m^*}\int_{2m^*}^{2m^*\xi}\left(1-\frac{2m^*}{r}\right)^{-\frac{1}{2}}dr$$

$$=(\xi^2-\xi)^{\frac{1}{2}}+\frac{1}{2}\ln\left[\frac{\xi^{\frac{1}{2}}+(\xi-1)^{\frac{1}{2}}}{\xi^{\frac{1}{2}}-(\xi-1)^{\frac{1}{2}}}\right]. \quad (44)$$

The radial coordinate, $r=2m^*\xi$, of the interface between two zone boundaries may be defined in two ways: (a) as the radius of the boundary sphere, or (b) as the radial coordinate of the midpoint between two mass particles, i.e., the center of the interface between two nonspherical zones. For Condition II both definitions are the same; the maximum radius is determined by

$$\xi_2=1/\sin^2\psi_2. \quad (45)$$

For Condition I one finds

$$\xi_1=1/\sin^2\psi_1, \quad (46a)$$

using the alternative (a), but a more complicated expression:

$$\xi_1\cot\psi_2=3\cot\psi_1+\cot^3\psi_1-2(\xi_1-1)^{\frac{1}{2}} \quad (46b)$$

with alternative (b).

With the rigorous metric (43), symmetry shows that a radial line drawn from one singularity at $r=\sqrt{3}+0.844$ to the Schwarzschild radius of the fifth particle at 11.314 will be a suitable geodesic. Therefore

$$2d=\int_{\sqrt{3}+0.844}^{11.314}\left(1+\frac{2\sqrt{2}}{r-\sqrt{3}}+\frac{6\sqrt{2}}{l}\right)^2dr \quad (47)$$

with

$$l^2=8/3+(r+\sqrt{3}/3)^2.$$

Evaluation of this integral gives

$$2d=83.70$$

or, in dimensionless units,

$$d/2m^*=0.925. \quad (48)$$

Values of this dimensionless ratio, as determined from the Schwarzschild cell approximation and from the rigorous metric, are given in Table V for the five and eight-body universes. Once again the Schwarzschild-cell treatment gives more reasonable results using Condition I, which allows the spheres of influence to overlap slightly.

Given the rigorous metric (43) at a reasonably fine lattice of points, an electronic computer can in principle calculate the metric at ensuing time intervals. Such a calculation may become practicable in the future. Recently progress has been made at Livermore, Los Alamos, and Princeton in treating problems of hydrodynamics where the velocity, pressure, and density depend on two space coordinates and one time coordinate. In the present problem the metric quantities depend upon one more space coordinate. It would be easier first to treat the problem of two mass singularities in an asymptotically flat universe. This problem idealizes the head on collision of two masses in a closed universe by assuming that the two masses in question are very far away from all the other masses that curve up the metric into a closed space.

TABLE V. Distances along a geodesic, at the moment of maximum expansion, from the Schwarzschild singularity to the boundary of a zone, in units of the Schwarzschild radius, $2m^*=2Gm/c^2$, of one of the masses.

Number of particles in lattice universe		5	8
Distance in Schwarzschild metric from singularity to spherical boundary, where radius of sphere is determined by Condition	I	1.187	1.812
	II	1.694	2.295
Distance in Schwarzschild metric from one singularity to point halfway to a nearest neighbor (center of interface between two zones), using Condition	I	0.961	1.599
	II	1.694	2.295
Distance in rigorous metric (43) from one singularity to point halfway to a nearest neighbor		0.925	1.609

7. POSSIBLE GENERALIZATIONS

Variational Formulation

Can the problem of a lattice universe be formulated in a way that does not make such heavy demands on the symmetry of the elementary lattice cell? An analogous familiar problem is to solve

$$\nabla^2\psi+k^2\psi=0, \quad (49)$$

in a two-dimensional region subject to the condition that ψ vanish on a boundary of unsymmetric shape, such as a triangle. One approximate method of solution uses the variation principle,

$$k^2=\text{minimum of }\frac{\int(\nabla\psi)^2dxdy}{\int\psi^2dxdy}. \quad (50)$$

For example, from the center of gravity of the triangle lines are drawn to the three corners to divide the triangle into three zones, I, II, III. In zone I a trial solution is defined by

$$\psi_{\text{trial I}}=\frac{\sin(k\cdot\text{distance from outer edge to }x,y)}{\sin(k\cdot\text{distance from outer edge to center})}; \quad (51)$$

similarly in II and III. The trial solution is everywhere continuous and satisfies the boundary conditions; it also satisfies the differential equation (49) everywhere except at the boundaries between zones. This permits one to express the right-hand side of (50) in terms of integrals taken exclusively along zone boundaries. In each such integral there enters the jump in the normal derivative of ψ across the boundary. These jump integrals are easily evaluated, giving a simple approximate eigenvalue for the wave equation in the region in question.

In analogy one can try to use the variational principle of general relativity,

$$\delta \int R(-g)^{\frac{1}{2}} dx^1 dx^2 dx^3 dx^4 = 0, \tag{52}$$

to formulate the Schwarzschild-cell method more precisely. The trial metric is constructed by joining together N-Schwarzschild metrics into an over-all metric of the appropriate symmetry. The boundaries are defined by the intersections of the Schwarzschild cells. They are not spherical. They are perhaps most easily visualized in the appropriate Euclidean imbedding space. The trial metric is uniquely determined by a single unknown function. It can be considered to represent the radius of the universe as a function of time, or the separation of nearest neighbors as a function of time. The variational integral (52) can be expressed in terms of the properties of this unknown function. The variational principle leads to a differential equation for its determination. This procedure gives the best possible Schwarzschild cell approximation to the accurate solution, "best" in the sense of the variational principle. We see no reason to expect from this method results substantially better than or different from those that come from a more intuitive formulation of the Schwarzschild cell method.

The dependent variables do not match smoothly at zone boundaries in the variational formulation of the problem, either of the lattice universe or of the wave equation (49). The variational principle minimizes the discrepancy only on the average, not locally. Therefore one cannot expect the spherical lattice cell approximation to satisfy the boundary conditions of O'Brien and Synge exactly.[12]

These authors deduced the boundary conditions that must be satisfied by the metric and its derivative at any three-dimensional surface of discontinuity in space-time. If this surface is defined by

$$x^4 = 0,$$

where the coordinate x^4 may be either space-like or time-like, the quantities

$$g_{\mu\nu},\ \frac{\partial g_{\mu\nu}}{\partial x^i},\ \frac{\partial^2 g_{\mu\nu}}{\partial x^i \partial x^j} \quad \text{and} \quad \frac{\partial g_{ij}}{\partial x^4} \begin{pmatrix} i,j=1,2,3 \\ \mu,\nu=1,2,3,4 \end{pmatrix} \tag{53}$$

must all be continuous. Thus one can only permit jumps in the four derivatives

$$\partial g_{4\mu}/\partial x^4. \tag{54}$$

The variational treatment gives up the continuity requirements on the quantities (53) locally, but satisfies them on the average.

Coalescence of Schwarzschild Singularities

What happens to the lattice universe at late stages of the contraction? The interface between two lattice cells is specified in the Schwarzschild-cell approximation by the motion of an infinitesimal test particle falling towards one Schwarzschild singularity; but it is well known that the test particle will arrive at the singularity in a finite proper time[13] [compare (30) or (32)] even though the necessary interval of Schwarzschild time is infinite. It will no longer be reasonable to approximate a lattice cell as a spherically symmetric region when the separation of two lattice cells becomes comparable to the radius of the Schwarzschild singularity at the center of each. From Table IV it appears that this degree of approach will be attained relatively soon in the case of the 5-mass system, and relatively late for the system of 600 mass centers. The surface of singularity will be deformed from its normal spherical form, $r=2m^*$. This deformation might trigger off an exponentially growing disturbance in the singularity, causing it to become violently disturbed and possibly to explode. Such an outcome seems unlikely in view of considerations in another paper on the stability of a Schwarzschild singularity.[14]

If stability obtains, three alternatives suggest themselves for the manner of approach of the singularities: (1) Approaching faces flatten more and more. The remaining space continuum has the character of a thin film separating polyhedral dice from each other. Ultimately this film is pressed to nothingness. (2) The film becomes ever thinner but never disappears. (3) Approaching faces bulge out to meet each other. On contact they join. The regions interior to the N-Schwarzschild singularities suddenly become connected to each other. Equidistant from neighboring singularities there are still regions of space. Moreover, there still remain channels between these regions. However, a little later these channels are pinched off and the regions of space become islands. The islands continue to shrink in size and eventually disappear.

To analyze by way of the field equations the limiting analytical forms for the metric near a point or surface

[12] S. O'Brien and J. L. Synge, *Jump Conditions at Discontinuities in General Relativity* (Dublin, Institute for Advanced Studies, 1952).

[13] See reference 8. Synge cites there H. P. Robertson's announcement of the same conclusion in 1939.

[14] T. Regge and J. A. Wheeler (to be published).

of coalescence is a fascinating problem for the future, the answer to which might help to decide between these alternatives or some other possible outcome.

When the disposition of the N-mass centers in the lattice universe is almost but not quite symmetric, a situation arises much like that in a Freidmann universe when the mass distribution departs slightly from uniformity. Unless the initial conditions are very special, the magnitude of the disturbance will grow. When underwater bubbles undergo dilatational oscillations, and when the surface departs slightly from spherical symmetry at the phase of maximum expansion, then the magnitude of the disturbances ordinarily grows. In this case one can follow the phenomena far enough visually to see that prongs and spikes form. The impression is gained that the bubble changes over from contraction to expansion, not everywhere simultaneously over its surface, but more after the fashion of a glove being turned inside out one finger at a time. If the analogy is any guide, the not quite symmetrical lattice universe will be expected to show a similar behavior. One will expect first a few Schwarzschild singularities to amalgamate and then break apart, then others to fall in, amalgamate, and break apart again, and so on, with some parts of the system therefore still contracting while others have already begun reexpansion. To show the beginnings of such a behavior, a perturbation theory analysis of the regular lattice universe should suffice. To follow the later and more interesting phases of the turnabout would demand a much more elaborate scheme of analysis.

ACKNOWLEDGMENTS

We are indebted to Charles Misner for many discussions of the problem of the lattice universe and for his help in constructing the special metric of Sec. 6.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Observables in Singular Theories by Systematic Approximation*†

EZRA NEWMAN‡ AND PETER G. BERGMANN

Syracuse University, Syracuse, New York

1. INTRODUCTION

IN any general-relativistic theory the field variables necessarily carry some information that relates to the choice of frame of reference rather than to the physical situation. Two manifestly different fields of gravitational potentials may describe one and the same gravitational field, merely in terms of two different coordinate systems. General relativity differs from special relativity in the degree of freedom inherent in the choice of coordinate system. Whereas in a Lorentz-covariant theory the frame of reference may be chosen at one instant in time and then remains fixed (the whole freedom of choice reducing to the determination of ten parameters once and for all), the freedom in a general-relativistic theory amounts to the determination of four arbitrary functions throughout space, anew at each instant in time. As a result of this vast freedom of choice, general-relativistic dynamical laws cannot be expected[1,2] to permit the integration of the field equations in the sense that suitable initial-value conditions at one time t_0 predict the value of any component of the metric tensor $g_{\mu\nu}$, or, for that matter, of any other conventional field variable, at some space point x^s at a different time t. Nevertheless, Einstein's general theory of relativity is quite deterministic. Its field equations do determine the gravitational field from initial-value conditions for all times to come. The metric potentials are simply not the quantities that are determined completely by the physical situation.

Failure of the mathematical theory to predict the value of a field variable at a given world point corresponds to physical unobservability. It is impossible to devise an experiment that will measure some field at the world point with the coordinates x^μ, because the values of the coordinates by themselves do not identify that world point. In actual practice, a world point at which some measurement is to be made is always identified in some other manner, such as the convergence of a beam of light or the location of a material component of our instrumentation. The determination of a gravitational potential at a world point and in directions that are defined by the values of specified electromagnetic quantities represents, of course, something different from the determination of that same potential at a world point specified by nothing but

* Work leading up to this paper was supported at various times by the Office of Naval Research, the National Science Foundation, and the Air Force Office of Scientific Research.

† This paper contains the results of a doctorate dissertation by one of us (E.N.), which has been accepted by the Graduate School of Syracuse University.

‡ Now at University of Pittsburgh, Pittsburgh, Pennsylvania.

[1] P. G. Bergmann, Phys. Rev. 75, 680 (1949).

[2] P. G. Bergmann and R. Schiller, Phys. Rev. 89, 4 (1953).

coordinate values. The theory may be expected to predict (from initial-value conditions) the former, but not the latter.

A quantity that can be predicted by the theory and that is not trivial (such as the value of a numeric) will be called an *observable*.[3] In a Lorentz-covariant theory all dynamical variables may be observables in this technical sense (though in electromagnetic theory only gauge-invariant quantities are observables); in a general-relativistic theory usually none of the ordinary fields are. But if a field variable cannot even be predicted in c-number theory, then surely it cannot have an expectation value in a quantized version of the theory, either. Hence the construction of observables in our sense is a necessary preliminary for quantizing any theory in which the original variables of the theory do not have that property.

If the dynamical laws of a theory have the form of partial differential equations and if the theory is to predict observables at a time t from suitable initial-value conditions at a time t_0, then knowledge of the values of a sufficient number of observables at any intermediate time t', between t_0 and t, must permit the same prediction. We shall call such a set of observables at a time t' that permits prediction at other times a *complete set of observables*. The construction of a complete set of observables is the central theme of this paper.

Mathematically, the failure of a general-relativistic theory to permit prediction of ordinary field variables by integration expresses itself in the "singular" character of its action principle. If, as is usually the case, the Lagrangian of the action principle is a function of the field variables themselves and of their first partial derivatives, the resulting field equations will be of the second order. If we now separate the derivatives in a particular direction (say with respect to x^4) from the remainder and focus attention on the second "time" derivatives, it turns out that the field equations cannot be solved with respect to them. Even though the number of field equations equals the number of field variables, (at least) four linear combinations of the second "time" derivatives are completely unrestricted by the field equations, whereas (at least) four independent combinations of the field equations contain no reference to the second "time" derivatives at all. In a canonical (Hamiltonian) formulation, "singular" action principles lead to Hamiltonians that are only partially determined and to constraining algebraic relations between the canonical field variables.[4] Exactly the same situation is also met with in electrodynamics, where it results from the arbitrariness of the gauge frame.

Complete sets of observables are fairly easily obtained in electrodynamics and in other linear and quasi-linear theories.[5,6] Because of its much greater complexity, the general theory of relativity has so far not permitted the construction of any observables, much less a complete set. In this paper we develop a systematic method of construction by means of successive approximations, the lowest nontrivial stage corresponding to the linearized theory. As in other methods of successive approximations in physics, the question of convergence remains unanswered. Likewise, we have developed our scheme as a procedure "in the small," leaving all questions of topology aside for the time being. What we have accomplished is to formulate a procedure that at every stage leads to solvable problems. Both the method as such and its products, variables that approximate observables, lend themselves to intuitive interpretation.

The strategy of our procedure is primarily adapted to a general-relativistic theory, or more particularly to Einstein's theory of gravitation (general relativity), but is applicable to any theory whose singular character is the result of invariance properties. We shall assume that there exists a known trivial solution of the field equations (corresponding to the Minkowski metric), which consists of a constant field. The expansion is in terms of deviation from this trivial standard solution. The lowest nontrivial equations are then the linearized field equations. They correspond to an action principle that is singular (in the sense used above) and invariant with respect to a transformation group that has some similarity with the group of curvilinear coordinate transformations. At each successive stage we construct a more complex Lagrangian (nth order Lagrangian) that is singular and invariant with respect to a transformation group that increasingly resembles the group of coordinate transformations. Solving the Euler-Lagrange equations of the nth order involves correcting the $(n-1)$st-order solutions by solving a linear problem. Likewise, once we have constructed the observables corresponding to the first-order problem, constructing the observables of the next stage requires corrections that are obtained from x linear conditions. All of our work will be carried out by means of Hamiltonian (rather than Lagrangian) techniques.

Because the notation gets quite involved, we present the method in terms of a problem involving a system with a finite number of degrees of freedom, in other words a contrived "mechanical" system. Also to simplify the presentation, we assume that only first-class constraints are present.

2. INVARIANCE PROPERTIES OF THE APPROXIMATE THEORY

We assume that the exact equations of motion are derivable from an action principle, with a known Lagrangian $L(q_i,\dot{q}_i)$. The coordinate variables are then written as a power expansion in some small parameter:

$$q_i=\sum_{a=0}^{\infty} \epsilon^a q_{(i\ a)}. \tag{2.1}$$

[3] P. G. Bergmann and I. Goldberg, Phys. Rev. 98, 531 (1955).
[4] P. A. M. Dirac, Can. J. Math. 2, 129 (1950); 3, 1 (1951).
[5] P. A. M. Dirac, Can. J. Phys. 33, 650 (1955).
[6] P. G. Bergmann, Nuovo cimento 3, 1177 (1956).

The zeroth-order variables $q_{i,0}$ are taken to be constant. They are, for example, the components of the Minkowski metric in gravitational theory. This method of approximation is very similar to the Einstein-Infeld-Hoffmann (EIH) method. The equations of motion for each variable of different order are linear nonhomogeneous equations. The nonhomogeneous part contains the variables of the previous orders. For convenience we shall call this hierarchy of equations the EIH equations. The immediate problem is to find a Lagrangian that will yield the EIH equations, to any order, as Euler equations. We shall show that when the full Lagrangian is expanded in powers of ϵ, the ath order Lagrangian will yield the first a EIH equations. To prove this write the total action

$$S=\int Ldt,$$

and vary the total q_i's. This gives

$$\delta S=\int \delta Ldt=\int L^i\delta q_i dt, \tag{2.2}$$

where L^i stands for the total Euler equations and the repeated index i denotes summation (Sums over indices denoting different orders will be indicated explicitly.) Since

$$L^i=\sum_{a=1}^{\infty}\epsilon^a L^i{}_a, \quad \delta q_i=\sum_{b=1}^{\infty}\epsilon^b\delta q_{(i,b)}, \tag{2.3}$$

where $L^i{}_a$ are the EIH equations, (2.2) yields

$$\delta S=\int\sum_{a=1}^{\infty}\sum_{b=1}^{\infty}\epsilon^{a+b}L^i{}_a\delta q_{(i,b)}$$
$$=\sum_{c=2}^{\infty}\epsilon^c\sum_{a=1}^{c-1}\int L^i{}_a\delta q_{(i,c-a)}dt. \tag{2.4}$$

Instead of first varying the action and substituting the expanded variables as in (2.2) and (2.3), substitute the expanded variables and then vary the action. If

$$L=\sum_{1}^{\infty}\epsilon^{a+1}L_a$$

(here we make the assumption that the Lagrangian is at least quadratic) then the action integral reads

$$S=\sum_{1}^{\infty}\epsilon^{a+1}\int L_a dt.$$

The variation of S gives

$$\delta S=\sum_{1}^{\infty}\epsilon^{a+1}\int\delta L_a dt=\sum_{a=1}^{\infty}\epsilon^{a+1}\int\sum_{b=1}^{a}L_a{}^{(i,b)}\delta q_{(i,b)}dt$$
$$=\sum_{a=2}^{\infty}\epsilon^a\sum_{b=1}^{a-1}\int L_{a-1}{}^{(i,b)}\delta q_{(i,b)}dt, \tag{2.5}$$

where $L_a{}^{(i,b)}$ are the Euler equations of the Lagrangian L_a, corresponding to the variable $q_{(i,b)}$. By comparing the coefficients of ϵ in (2.4) and (2.5), we have the identity

$$\sum_{b=1}^{a-1}\int L^i{}_b\delta q_{(i,a-b)}dt=\sum_{b=1}^{a-1}\int L_{a-1}{}^{(i,b)}\delta q_{(i,b)}dt. \tag{2.6}$$

Since the variations in the $q_{(i,b)}$'s are to be taken independently, the coefficients in the integral must be identical giving

$$L^i{}_{a-b}\equiv L_a{}^{(i,b+1)}\equiv L_{a+d}{}^{(i,b+d+1)}. \tag{2.7}$$

This means that the $(a-b)$th EIH equation is equal to the $(b+1)$st Euler equation of the ath Lagrangian, L_a. For a fixed a we can get the first a EIH equations by letting b go from zero to $a-1$.

According to E. Noether's theorem, a theory which possesses an invariance group depending on arbitrary functions also possesses differential identities among the Euler equations, which we may call the generalized Bianchi identities. We derive the generalized Bianchi identities with an assumed transformation law, and by expanding them in a power series, show that the Euler equations of the ath Lagrangian possesses Bianchi identities as well. It will also be necessary to derive the transformation law under which the ath Lagrangian is invariant.

Assume the total Lagrangian to be invariant under the infinitesimal transformation

$$\bar{\delta}q_i=F_i{}^jq_j\dot{\xi}-\dot{q}_i\xi. \tag{2.8}$$

The bar distinguishes an infinitesimal transformation from an arbitrary variation of q_i. This transformation is the analog of the gauge and coordinate transformations in electrodynamics and relativity, respectively. We get the Bianchi identities by noting that[7]

$$\bar{\delta}L\equiv\dot{Q}\equiv\frac{\partial L}{\partial q_i}\bar{\delta}q_i+\frac{\partial L}{\partial\dot{q}_i}\bar{\delta}\dot{q}_i\equiv L^i\bar{\delta}q_i+\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{q}_i}\bar{\delta}q_i\right). \tag{2.9}$$

Substituting (2.8) into (2.9) and differentiating once by parts we obtain

$$\left[\dot{q}_iL^i+\frac{d}{dt}(F_i{}^jL^iq_j)\right]\xi\equiv\frac{d}{dt}\left[\frac{\partial L}{\partial\dot{q}_i}\bar{\delta}q_i-Q+F_i{}^jq_jL^i\xi\right]. \tag{2.10}$$

Since the ξ's are completely arbitrary, we have

$$\frac{d}{dt}(L^iF_i{}^jg_j)+L^i\dot{q}_i\equiv 0, \tag{2.11}$$

the Bianchi identities of the full theory. The power series of these identities is obtained by straight sub-

[7] P. G. Bergmann and R. Schiller, Phys. Rev. 89, 4 (1953).

stitution from (2.3)

$$F_i{}^j\frac{d}{dt}(L^i{}_1q_{(j,0)})\equiv 0,$$
$$F_i{}^j\sum_{a=1}^{c}\frac{d}{dt}(L^i{}_aq_{(j,c-a)})+\sum_{a=1}^{c-1}L^i{}_a\dot{q}_{(i,c-a)}\equiv 0,\quad c\geqslant 2. \tag{2.12}$$

To discuss the Bianchi identities of the approximate theory, we derive the transformation law for the $q_{(i,a)}$ of different order. By writing

$$\xi=\sum_{a=0}^{\infty}\epsilon^a\xi_a, \tag{2.13}$$

and by replacing q_i and $\dot{q}_i$ by their power expansion in (2.8), we get

$$\bar{\delta}q_i=\sum_{a=1}^{\infty}\epsilon^a\bar{\delta}q_{(i,a)}$$
$$=F_i{}^j\sum_{a=0}^{\infty}\epsilon^aq_{(j,a)}\sum_{b=0}^{\infty}\epsilon^b\dot{\xi}_b-\sum_{a=1}^{\infty}\sum_{b=0}^{\infty}\epsilon^{a+b}\dot{q}_{(i,a)}\xi_b. \tag{2.14}$$

Reordering terms in the sums and equating the different coefficients of ϵ^a we obtain[8]

$$\bar{\delta}q_{(i,a)}=F_i{}^j\sum_{b=0}^{a}q_{(j,a-b)}\dot{\xi}_b-\sum_{b=0}^{a-1}\dot{q}_{(i,a-b)}\xi_b. \tag{2.15}$$

The requirement that $q_{(i,0)}$ equal a fixed constant results in ξ_0 being a constant. In general relativity, likewise, the $g_{(\mu\nu,0)}$ will remain Minkowskian only if the zeroth-order coordinate transformations are Lorentz transformations.

We can establish whether the transformation law (2.15) will leave a particular Lagrangian L_b invariant. To do so we assume that the Lagrangian *is* in fact invariant under the law (2.15), and then show that the resulting Bianchi "identities" are actually identically satisfied by virtue of (2.12).[9]

We perform the identical operations that we used to obtain (2.11), except that we replace L by L_a, and the transformation law (2.8) by (2.15):

$$\sum_{b=1}^{a}\sum_{c=1}^{b}F_i{}^j\frac{d}{dt}(L_a{}^{(i,b)}q_{(j,b-c)})\xi_c+\sum_{b=1}^{a}\sum_{c=1}^{b-1}L_a{}^{(i,b)}\dot{q}_{(i,b-c)}\xi_c\equiv 0. \tag{2.16}$$

If we now interchange the sums, being careful to use the correct new ranges of summation, then because the ξ_c's are arbitrary, we obtain

$$F_i{}^j\frac{d}{dt}(L_a{}^{(i,a)})q_{(j,0)}\equiv 0,$$
$$\sum_{b=c}^{a}F_i{}^j\frac{d}{dt}(L_a{}^{(i,b)}q_{(j,b-c)})+\sum_{b=c+1}^{a}L_a{}^{(i,b)}\dot{q}_{(i,b-c)}\equiv 0. \tag{2.17}$$

After use of (2.7) and some relabeling of the indices, (2.17) reduces exactly to (2.12). This completes the proof that the approximate theory possesses the same invariance properties as the full theory to any order in the approximation. Although in the full theory there is only one Bianchi identity for each ξ, in the "truncated" theory to order a, there are a identities. The reason for this is that the transformation group now depends on the a arbitrary functions, ξ_a, instead of just one.

3. ABSENCE OF FALSE CONSTRAINTS

The matrix $L^{ij}\equiv\partial^2L/\partial\dot{q}_i\partial\dot{q}_j$ is singular in all theories that possess an invariance group depending on arbitrary functions; the converse, however, is not true. L^{ij} may be singular and possess no invariance group; in that case the singularity corresponds to second-class constraints (constraints having nonvanishing Poisson brackets with each other) in the canonical formalism. With our assumption of only one arbitrary function in the invariant transformation law and the existence of only first-class constraints, the equation $L^{ij}u_i=0$ has only one independent solution and there are exactly two first-class constraints.[10] The first constraint (called *primary* constraint) expresses the fact that the velocities are not uniquely determined by the q's and p's. The second constraint (*secondary* constraint) arises from the consistency requirement that the time derivative of the primary constraint must vanish. The matrix corresponding to L^{ij} in the approximate theory is

$$L_e{}^{(i,a)(j,b)}\equiv\frac{\partial^2L_e}{\partial\dot{q}_{(i,a)}\partial\dot{q}_{(j,b)}}.$$

Since the Lagrangian L_e is invariant under e transformations, the equation

$$L_e{}^{(i,a)(j,b)}u_{(i,a)}=0 \tag{3.1}$$

must have at least e solutions. The question arises as to whether the approximation procedure introduces additional solutions to (3.1), that is, whether it introduces second-class constraints. We will prove that this is not the case.

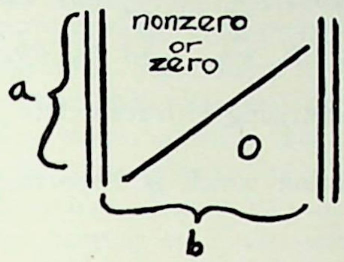
Because the original Lagrangian is assumed to be quadratic in the velocities, and because the expansion

[8] By showing that the Jacobi identity is satisfied, the group property of the transformation (2.15) can be verified.

[9] The fact that the Bianchi identities imply the invariance of the theory can be seen by replacing $\dot{Q}$ by F in Eq. (2.9). Continuing in the same manner as in the text, we get an equation similar to (2.11). If the quantities called the Bianchi identities *are* identically zero, the F must be equal to $\dot{Q}$, which is the condition for an invariant transformation.

[10] Bergmann, Janis, Goldberg, and Newman, Phys. Rev. 103, 807 (1956).

was written $L=\sum \epsilon^{a+1}L_a$, the matrix $L_e^{(i,a)(j,b)}$ takes the form



(3.2)

where every element stands for an $n\times n$ matrix, $i, j=1\cdots n$. Accordingly, (3.1) can be written

$$\sum_{d=1}^{c} L_e^{(j,e-c+1)(i,d)}u_{(i,d)}=0, \quad c=1\cdots e. \tag{3.3}$$

We now express $L^{ij}u_i$ in a power series in ϵ, and equate the coefficients to zero. These coefficients will equal (3.3); because of the uniqueness of the power expansion of u_i. Before we can perform this derivation, we need a lemma, which is also important in the next section.

If we consider x as a general variable, standing for q, $\dot{q}$, or p and write $x=\sum \epsilon^a x_a$, then any function of x can be expressed as a power series in ϵ, $f(x)=\sum \epsilon^a f_a(x_b)$. Our lemma can be stated in the form

$$\frac{\partial f_e}{\partial x_{e-d}}\equiv\frac{\partial f_{b+d}}{\partial x_b}\equiv\frac{\partial f_{c+d}}{\partial x_c}. \tag{3.4}$$

The proof is simple. We write $f(x)=f(\sum \epsilon^a x_a)$ and note

$$\frac{\partial f}{\partial x}=\epsilon^{-b}\frac{\partial f}{\partial x_b}=\epsilon^{-c}\frac{\partial f}{\partial x_c}. \tag{3.5}$$

The indicated differentiation operations are then performed on the expanded $f(x)$, and the coefficients of different powers of ϵ are compared; hence (3.4).

We are now ready to expand the equation $L^{ij}u_j=0$. If

$$u_i=\sum_{a=0}^{\infty} \epsilon^a u_{(i,a+1)} \tag{3.6}$$

(the form of the summation is just for convenience) and

$$L^{ij}=\sum_{a=0}^{\infty} \epsilon^a \frac{\partial^2 L_{a+b+c-1}}{\partial\dot{q}_{(i,b)}\partial\dot{q}_{(j,c)}} =\sum_{a=0}^{\infty} \epsilon^a L_{a+b+c-1}{}^{(i,b)(j,c)}, \quad b, c \text{ arbitrary}, \tag{3.7}$$

where we have used the power expansion of L and (3.5) twice, then

$$L^{ij}u_i=\sum_{d=0}^{\infty}\sum_{a=0}^{\infty} \epsilon^{a+d} L_{d+b+c-1}{}^{(i,b)(j,c)}u_{(i,a+1)} =\sum_{f=0}^{\infty} \epsilon^f \sum_{a=0}^{f} L_{f-a+b+c-1}{}^{(i,b)(j,c)}u_{(i,a+1)}. \tag{3.8}$$

After equating the coefficients of ϵ to zero and using (3.4) in the following form

$$\frac{\partial^2 L_{f-a+b+c-1}}{\partial\dot{q}_{(i,b)}\partial\dot{q}_{(j,c)}}\equiv\frac{\partial^2 L_e}{\partial\dot{q}_{(i,b)}\partial\dot{q}_{(j,e-f+a-b+1)}}\equiv L_e{}^{(i,b)(j,e-f+a-b+1)},$$

we obtain

$$\sum_{a=0}^{f} L_e{}^{(i,b)(j,e-f+a-b+1)}u_{(i,a+1)}=0. \tag{3.9}$$

By relabeling the indices and setting b equal to $a+1$, we obtain exactly (3.3). The arbitrariness in the choice of b yields, for example, the following e linearly independent null vectors of L

$$\left\|\begin{matrix} u_{(i,1)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_{(i,e)} \end{matrix}\right\|, \quad \left\|\begin{matrix} 0 \\ u_{(i,1)} \\ \cdot \\ \cdot \\ \cdot \\ u_{(i,e-1)} \end{matrix}\right\|, \cdots \left\|\begin{matrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_{(i,1)} \end{matrix}\right\|. \tag{3.10}$$

The matrix (3.2), being of order e can have at most e independent null vectors; if there were other null vectors, they would have to come from the n by n submatrices. Since the power expansion of L is unique and there is only one u_i which is a null vector, there can be only one of each type of the e vectors (3.10). This proves that there are only e null vectors of $L_e^{(i,a)(j,b)}$ and since the most general transformation of the "truncated" theory to order e contains e arbitrary functions, there can be no second class constraints.

4. CANONICAL FORMALISM

Construction of a Hamiltonian formalism from the Lagrangian L_e proceeds along lines indicated by Dirac.[4] It differs from the usual procedure in that the Hamiltonian is not a unique function of the canonical variables; a linear combination of primary constraints can be added. Though construction of the Hamiltonian is easy, it must be shown what relationship it bears to the expansion of the complete Hamiltonian formalism.

First write the expansion of the exact p_i in the form

$$p_i=\sum_{1}^{\infty} \epsilon^a \bar{p}_{(i,a)}, \tag{4.1}$$

and try to connect the $\bar{p}_{(i,a)}$'s with the

$$p_{e(i,a)}=\frac{\partial L_e}{\partial\dot{q}_{(i,a)}}.$$

This can be done as follows

$$p_i=\frac{\partial L}{\partial\dot{q}_i}=\sum_{a=b}^{\infty} \epsilon^{a-b+1}\frac{\partial L_a}{\partial\dot{q}_{(i,b)}}=\sum_{c=1}^{\infty} \epsilon^c \frac{\partial L_{b+c-1}}{\partial\dot{q}_{(i,b)}} =\sum_{c=1}^{\infty} \epsilon^c \frac{\partial L_e}{\partial\dot{q}_{(i,e-c+1)}}=\sum_{c=1}^{\infty} \epsilon^c p_{e(i,e-c+1)}, \tag{4.2}$$

using (3.4) and (3.5). Hence from (4.1) and (4.2)

$$\bar{p}_{(i,a)} = p_{e(i,e-a+1)}. \tag{4.3}$$

In the power expansion of any function $A(q,p)$, the $\bar{p}_{(i,a)}$'s must be replaced with the aid of (4.3) in order to obtain the variable or variables in the approximate theory which are equivalent to the A in the full theory. It is easy to show that the eth term in the Hamiltonian constructed in this manner is equal to the Hamiltonian formed from the eth Lagrangian,

$$H_e = \sum_{a=1}^{e} p_{e(i,a)} \dot{q}_{(i,a)} - L_e. \tag{4.4}$$

We must obtain a connection between the Poisson bracket of two functions in the full theory and the Poisson bracket in the approximate theory. This relationship may be stated in its simplest form by the equation

$$(A,B) = \sum_{1}^{\infty} \epsilon^{e-1} (A_e, B_e)^e, \tag{4.5}$$

where the bracket with the superscript e means the Poisson bracket of A_e with B_e using the variables $q_{(i,1)} \cdots q_{(i,e)}$ and $p_{e(i,1)} \cdots p_{e(i,e)}$. In the proof of (4.5) we suppress the coordinate index i.

Using (3.5), we can write

$$\begin{aligned} \frac{\partial A}{\partial q} &= \sum_{a=b}^{\infty} \epsilon^{a-b} \frac{\partial A_a}{\partial q_b} = \sum_{e=0}^{\infty} \epsilon^{e} \frac{\partial A_{b+e}}{\partial q_b}, \\ \frac{\partial B}{\partial p} &= \sum_{c=d}^{\infty} \epsilon^{c-d} \frac{\partial B_c}{\partial \bar{p}_d} = \sum_{f=0}^{\infty} \epsilon^{f} \frac{\partial B_{d+f}}{\partial \bar{p}_d}, \end{aligned} \tag{4.6}$$

from which follows

$$\frac{\partial A}{\partial q}\frac{\partial B}{\partial p} = \sum_{g=0}^{\infty} \epsilon^{g} \sum_{e=0}^{g} \frac{\partial A_{b+e}}{\partial q_b} \frac{\partial B_{d+g-e}}{\partial \bar{p}_d}. \tag{4.7}$$

With the aid of (3.4), this can be rewritten

$$\frac{\partial A}{\partial q}\frac{\partial B}{\partial p} = \sum_{g=0}^{\infty} \epsilon^{g} \sum_{e=0}^{g} \frac{\partial A_b}{\partial q_{b-e}} \frac{\partial B_d}{\partial q_{d-g+e}}. \tag{4.8}$$

The $\bar{p}$'s can be eliminated by the equation $\bar{p}_{d-g+e} = p_{a(a-d+g-e+1)}$, obtained from (4.3), yielding

$$\frac{\partial A}{\partial q}\frac{\partial B}{\partial p} = \sum_{g=0}^{\infty} \epsilon^{g} \sum_{e=0}^{g} \frac{\partial A_b}{\partial q_{b-e}} \frac{\partial B_d}{\partial p_{a(a-d+g-e+1)}}. \tag{4.9}$$

Since the b, d, and a are arbitrary, we choose them to be equal to $g+1$, with the result

$$\frac{\partial A}{\partial q}\frac{\partial B}{\partial p} = \sum_{g=0}^{\infty} \sum_{e=0}^{g} \frac{\partial A_{g+1}}{\partial q_{g+1-e}} \frac{\partial B_{g+1}}{\partial p_{g+1(g+1-e)}}. \tag{4.10}$$

By relabeling the indices we obtain

$$\frac{\partial A}{\partial q}\frac{\partial B}{\partial p} = \sum_{a=1}^{\infty} \epsilon^{a-1} \sum_{d=1}^{a} \frac{\partial A_a}{\partial q_d} \frac{\partial B_a}{\partial p_{a(d)}}. \tag{4.11}$$

After antisymmetrizing with respect to A and B, (4.5) results.

Another relation which is important in the next section is

$$(A_a, B_b)^c \equiv (A_a, B_{b+1})^{c+1} \equiv (A_{a+1}, B_b)^{c+1}, \quad c \geqslant a, b. \tag{4.12}$$

This can be proven simply by a direct calculation. It must be emphasized that the coefficients of the expanded functions of the q's and p's, though unique functions of the q_a's and $\bar{p}_a$'s, are not unique functions of the unbarred $p_{e(a)}$'s; the subscript e being arbitrary; (4.12) is to be understood as an identity in the $\bar{p}_a$'s.

The consistency of our approximation method with the full Hamiltonian theory can be tested. From the Poisson bracket $(q,p)=1$, it should be possible to derive the relations $(q_a, p_{e(b)})^e = \delta_{ab}$. That this is true can be seen from (4.5) and (4.12). When A and B are replaced by q and p in (4.5), we get the result

$$\begin{aligned} (q_1, \bar{p}_1)^1 &= (q_1, p_{1(1)})^1 = 1, \\ (q_e, \bar{p}_e)^e &= (q_e, p_{e(1)})^e = 0, \end{aligned} \tag{4.13}$$

where we have used (4.3) to eliminate the barred p's. With the aid of (4.12) applied to (4.13) we can obtain the entire system of Poisson bracket relations.

5. INVARIANTS IN THE APPROXIMATE THEORY

When a classical theory contains a general invariance group (depending on arbitrary functions rather than arbitrary parameters), the velocities cannot be solved as unique functions of the canonical variables from the equation

$$p_i = \frac{\partial L(q,\dot{q})}{\partial \dot{q}_i}. \tag{5.1}$$

The velocities can be eliminated from some of them yielding relationships between the canonical variables that must be satisfied if the Hamiltonian formalism is to be equivalent to the Lagrangian formalism. These relations, called primary first-class constraints, are denoted by $C_s(q,p)=0$. The number of these constraints is equal to the number of arbitrary functions in the invariant transformation law. For consistency, the time derivative of these constraints must vanish when the constraints themselves are satisfied. If this does not happen automatically then additional constraints must be imposed, $(C_s,H)=0$. These are called *first-class* secondary constraints. It is assumed here that no tertiary constraints appear. When we deal with the eth-order Lagrangian L_e, e primary constraints and e secondary constraints (if the full theory possesses one primary constraint) will appear.

According to Bergmann and Goldberg[3] the canonical variables are not the most appropriate variables for constructing the Hamiltonian formalism and transformation theory; the "observables" are the (maximum number of) independent functions of the canonical variables that have vanishing Poisson brackets with the constraints. They do not change their value or form under the invariant transformations of the theory. All quantities of physical interest can be expressed as functions of the observables. Their number in a given theory can be calculated readily to be $(2n-2n_0)$, $2n$ being the number of canonical variables and n_0 the number of constraints (this is not true if some of the constraints are second-class). In our case, due to our assumptions, the full theory possesses $(2n-2)$ observables and the eth approximation $(2n-2)e$.

The purpose of this section is to show how to construct the $(e+1)$th observables of the $(e+1)$-st approximation if we know the e observables of the eth approximation. If we write the observables as $O_a{}^u$, a going from 1 to e and u going from 1 to $(2n-2)$ and the constraints $C_a{}^s$, s going from 1 to 2, we shall show that the O_a's are the power expansion of the full O's. By replacing A and B in (4.5) by O^u and C^s and remembering that the full observables and the constraints must commute, we obtain

$$(O_a{}^u, C_a{}^s)^a = 0, \quad a = 1 \cdots e. \tag{5.2}$$

Repeated application of the identity (4.12) to (5.2) yields

$$(O_a{}^u, C_b{}^s)^e = 0. \tag{5.3}$$

This is just the condition for the $O_a{}^u$'s to be observables.

The first e observables of the $(e+1)$ approximation are just the e observables that we already know, the only difference being that all $p_{e(a)}$ are to be replaced by $p_{e+1(a+1)}$. That they commute with the constraint $C_{e+1}{}^s$ is proven by applying (4.12) once to (5.3) and taking $b=e$.

The problem that remains to be solved is how to construct the $(e+1)$st observables. In what follows we consider all functions as functions of the $\bar{p}$'s to simplify the notation. From (3.4)

$$\frac{\partial O_{e+1}{}^u}{\partial q_{(i,\,a+1)}} = \frac{\partial O_e{}^u}{\partial q_{(i,\,a)}}, \quad \frac{\partial O_{e+1}{}^u}{\partial \bar{p}_{(i,\,a+1)}} = \frac{\partial O_e{}^u}{\partial \bar{p}_{(i,\,a)}}. \tag{5.4}$$

These expressions indicate that the dependence of $O_{e+1}{}^u$ on all the q_a's and $\bar{p}_a$'s except for $a=1$, is uniquely determined. In other words, we now determine a function of $q_{(i,\,1)}$, and $\bar{p}_{(i,\,1)}$, which must be added to the already determined part of $O_{e+1}{}^u$. We can narrow the possibilities further by noting that $O_{e+1}{}^u$ are the coefficients of ϵ^{e+1} in a power series, and hence the function we are seeking must be homogeneous of the $(e+1)$st degree in the $q_{(i,\,1)}$ and $\bar{p}_{(i,\,1)}$. Thus we see that only the numerical coefficients are not determined. They, however, may be determined by the requirement

$$(O_{e+1}{}^u, C_a{}^s)^{e+1} = 0. \tag{5.5}$$

6. CONCLUSION

In the present paper we have developed an approximation procedure that purports to lead to the solution of the following purely abstract problem. Given a field y_A of several parameters x, with an (infinitesimal) transformation law under a group of transformations whose elements depend on one or several arbitrary functions of the x; find a complete set of functions (or functionals) of the $y_A(x)$ that are invariant with respect to the transformation group. An invariant, in this sense, differs from a scalar field in that its value is to remain unchanged if the transformed functions of x, $y_A(x)$ are substituted into its law of formation instead of the original ones, whereas a scalar retains its numerical value at a fixed point P (whose coordinates x will change their values as a result of the transformation). To the extent that the law of formation of an invariant makes any reference to the values of the parameters x at all, the same numerical values are to be inserted in both cases.

If the x are invariant under the group of transformations (as they are, e.g., in the gauge group), invariant and scalar are equivalent. But if the transformation group involves coordinate transformations in x space, they are not.

In this paper we have developed our approximation procedure within the Hamiltonian formalism. Originally it had been hoped that the approximation procedure would lead directly to a prescription for quantization, so that one could also speak of a quantum theory correct up to the ath order. Our discussion of the Poisson bracket in the c-number theory shows that such a procedure is not automatic. We shall endeavor to develop such a theory, either with the help of our approximation procedure in the canonical formalism or by Lagrangian quantization.

The motivation of this paper is closely related to work by Komar,[11] Géhéniau,[12] and Janis.[13] Komar and Géhéniau have proposed to construct invariants with the help of ideas specific to Riemannian geometry. Their ideas appear to apply primarily to pure gravitation theory, where they may be particularly useful; they do not represent a general physically oriented search for observables in singular theories. Janis, on the other hand, proposes to exploit a generalized Fermi-type approach involving subsidiary conditions, whose innocuous character is explicitly demonstrated.

[11] Ph.D. thesis, Princeton University, Princeton, New Jersey.
[12] Géhéniau, Helv. Phys. Acta Suppl. (to be published).
[13] A. Janis, Bull. Am. Phys. Soc. Ser. II, 2, 12 (1957).

Transformation Groups in a Q-Number Configuration Space*

IRWIN GOLDBERG

Department of Physics, Syracuse University, Syracuse, New York

THE quantum theoretical reformulation of classical theories frequently meets with difficulties that arise from ambiguities in the proper order of non-commuting factors as well as from the cumbersome transformation properties of canonical momentum components in theories that are to be relativistically invariant or gauge invariant. A number of authors have dealt with the possibility of circumventing some of these difficulties by basing the quantization procedure on the Lagrangian rather than the Hamiltonian version of a classical theory.[1–4]

However, not all quantum effects, notably spin, are derivable from the quantization of a classical theory. In these circumstances, it is worthwhile to seek a method for determining the observables, their commutators and their equations of motion of a quantum theory based on a q-number Lagrangian. In this paper a program is discussed for extending a classical transformation group in configuration space[5] to a q-number configuration space. The transformation group is to be related to the unitary transformations in Hilbert space. The variational principle approach to the equations of motion leads to difficulties because if the variations of the configuration variables are q numbers they cannot be separated from the Lagrangian terms unless their commutators with all the field variables are known, and if the variations are c numbers the conservation laws which are associated with q-number changes in the configuration variables are not easily determined. We shall postulate as the dynamical principle the isomorphism between the transformation group in configuration space and the unitary transformations in Hilbert space.

The infinitesimal unitary transformation of the operator C_1 is given by

$$C_1'-C_1=\bar{\delta}C_1=\exp\left(\frac{i}{\hbar}\epsilon C_2\right)C_1\exp\left(-\frac{i}{\hbar}\epsilon C_2\right)-C_1$$

$$=\epsilon\frac{i}{\hbar}[C_2,C_1]. \quad (1)$$

ϵ is an arbitrarily small constant, C_2 is a Hermitian operator called the generator of the transformation, and $[C_2,C_1]$ is the commutator of C_1 and C_2. The generator of the commutator of two infinitesimal unitary transformations is given by

$$C_{12}=\frac{i}{\hbar}\epsilon[C_1,C_2]. \quad (2)$$

We have attempted to set up a q-number transformation group which may be connected with the unitary transformations. To preserve the physical content of the theory the transformations must preserve the action. Then the Lagrangian L can at most change by a total time derivative.

$$\bar{\delta}L=L'(q',\dot{q}')-L(q,\dot{q})=\dot{Q}, \quad (3)$$

where the primed quantities are the result of the transformation. As in classical theory the functional dependence of the Lagrangian on the velocities may result in the singularity of the matrix $L_{ij}=\partial^2L/\partial\dot{q}^i\partial\dot{q}^j$. This singularity is usually the result of invariance of the theory under certain transformations such as gauge transformations in electrodynamics, or to some other structural properties of the Lagrangian.

In the canonical formulation of a classical theory the singularity of the L_{ij} indicates that not all of the velocities can be expressed as unique functions of the coordinates and momenta. This leads to restrictions on the canonical variables called constraints. Certain functions of the canonical variables generate transformations which alter the constraints. In the quantized theory the constraints become zero operators, and the correspondence principle between the canonical transformations and the unitary mappings leads to nonvanishing commutators between canonical variables and constraints. To avoid similar contradictions in the Lagrangian formalism we restrict our transformations to those mappings which preserve the functional form of the Lagrangian. Invariance properties and constraints arise from the form of dependence of the Lagrangian on the coordinates and velocities. If there is no change in the form of the Lagrangian, no change of invariance properties or constraints is possible. This restriction to invariant transformations is given by

$$\delta'L=L'(q,\dot{q})-L(q,\dot{q})=\dot{Q}-\frac{\partial L}{\partial q^i}\bar{\delta}q^i-\frac{\partial L}{\partial \dot{q}^i}\bar{\delta}\dot{q}^i=0. \quad (4)$$

---

* From a talk delivered at the Conference on Gravitation at Chapel Hill, North Carolina, January, 1957.

1 R. P. Feynman, Revs. Modern Phys. 20, 367 (1948).

2 J. Schwinger, Phys. Rev. 82, 914 (1951).

3 R. E. Peierls, Proc. Roy. Soc. (London) A214, 143 (1952).

4 P. G. Bergmann and R. Schiller, Phys. Rev. 89, 4 (1953).

5 Bergmann, Goldberg, Janis, and Newman, Phys. Rev. 103, 807 (1956).

n order to preserve the factor sequences any repeated ıdices will imply a "symbolic product." The term :moved by differentiation is replaced by the multiplied erm with the same index. In Eq. (4), for example, δq^j replaces the term removed from L by differentiation. We rewrite (4) by adding and subtracting $L_k\bar{\delta}q^k$, where L_k is the Euler-Lagrange function $\partial L/\partial q^k - d/dt(\partial L/\partial\dot{q}^k)$. Equation (4) may then be written

$$\dot{C}+L_k\bar{\delta}q^k, \tag{5}$$

$C=(\partial L/\partial\dot{q}^k)\bar{\delta}q^k-Q$ is the generator of the transformation. The generator of the commutator of the transformations generated by C_1 and C_2 is obtained by substituting the commutator transformation

$$\bar{\delta}_{12}q^k=\frac{\partial\bar{\delta}_1q^k}{\partial q^j}\bar{\delta}q^j+\frac{\partial\bar{\delta}_1q^k}{\partial\dot{q}^j}\bar{\delta}_2\dot{q}^j-\frac{\partial\bar{\delta}_2q^k}{\partial q^j}\bar{\delta}_1q^j-\frac{\partial\bar{\delta}_2q^k}{\partial\dot{q}^j}\bar{\delta}\dot{q}^j$$

in Eq. (5). Then

$$\dot{C}_{12}=\frac{d}{dt}\{C_1,C_2\}=\frac{d}{dt}\{\bar{\delta}_1C_2-\bar{\delta}_2C_1+\bar{\delta}_1p_k\bar{\delta}_2q^k-\bar{\delta}_2p_k\bar{\delta}_1q^k$$

$$\bar{\delta}_1C_2=\frac{\partial C_2}{\partial q^j}\bar{\delta}_1q^j+\frac{\partial C_2}{\partial\dot{q}^j}\bar{\delta}_1\dot{q}^j,\quad p_k=\frac{\partial L}{\partial\dot{q}^k}. \tag{6}$$

If terms containing second time derivatives are separated from (5) they must vanish independently.

$$\left(\frac{\partial C}{\partial\dot{q}^j}-L_{jk}\delta^kq\right)\ddot{q}^j=0,\quad \frac{\partial C}{\partial\dot{q}^j}=L_{jk}\bar{\delta}q^k. \tag{7}$$

'e now employ (7) to modify the expression for $\bar{\delta}_1C_2$ a… or… by…

$$\bar{\delta}_1C_2=\frac{\partial C_2}{\partial q^j}\bar{\delta}_1q^j+L_{jk}\bar{\delta}_1q^k\bar{\delta}_2\dot{q}^j, \tag{8}$$

$_{jk}\bar{\delta}_2\dot{q}^j$ is replaced from the time derivative of (7)[5] and

$$\bar{\delta}_1C_2=\{C_1,C_2\}+L_k\left(\frac{\partial\bar{\delta}_1q^k}{\partial\dot{q}^j}\right)\bar{\delta}_2q^j. \tag{9}$$

If we adopt as our dynamical principle the equivalence between the unitary transformations and the invariant transformations in order that (1) be satisfied the field equations must require:

$$L_k\left(\frac{\partial\bar{\delta}_1q^k}{\partial\dot{q}^j}\right)\bar{\delta}_2q^j=0. \tag{10}$$

Equations (10) may be employed as the field equations, but there are too many equations and no apparent connection with the usual quantum equations of motion.

The connection between (10) and the Heisenberg equations may be discovered by studying the invariant transformation $\bar{\delta}q^k=\epsilon\dot{q}^k$. This transformation is generated by the infinitesimal Hamiltonian function

$$\epsilon H=\epsilon\left(\frac{\partial L}{\partial\dot{q}^k}\dot{q}^k-L\right). \tag{11}$$

The transformation generated by H may also be written

$$\bar{\delta}_HC=\epsilon\dot{C}-\frac{\partial C}{\partial t}. \tag{12}$$

We use (12) to rewrite (5) and obtain

$$\frac{1}{\epsilon}\{C,H\}+\frac{\partial C}{\partial t}=0. \tag{13}$$

The Heisenberg equations require

$$\frac{i}{\hbar}[C,H]+\frac{\partial C}{\partial t}=\dot{C}. \tag{14}$$

If the invariant transformations are to provide a Hilbert algebra consistent with the Heisenberg equations, the field equations must lead to

$$\dot{C}=0,$$

$$L_k\bar{\delta}q^k=0,$$

$$\frac{i}{\hbar}\epsilon[C_1,C_2]=\{C_1,C_2\}. \tag{15}$$

If (10) are satisfied (15) are satisfied as may be seen by setting $\bar{\delta}_1q^k=\epsilon\dot{q}^k$ in (10). One may demonstrate the complete equivalence of (10) and (15) by a more detailed calculation. For nonsingular theories the above procedure leads to a quantum formulation completely equivalent to the one obtained by the usual procedures.

For singular theories we are not yet convinced of the validity of our dynamical principle. This may be illustrated by a consideration of the classical problem. In a classical theory the singularity of the matrix L_{ij} leads to field equations which do not contain accelerations. If $L_{ik}U_{(s)}{}^k=0$ defines the sth null vector (eigenvector corresponding to zero eigenvalue) of L_{ik} the combinations of field equations $L_kU_{(s)}{}^k$ are relations between the coordinates and velocities, or constraints. The constraints reduce the number of degrees of freedom of the system. Correspondingly in the q-number theory the singularity should reduce the number of degrees of freedom of the system. This means that the number of independent invariant transformations is less than the number of coordinates and velocities, or equivalently, that the eigenvectors of the generators of invariant transformations do not span the linear vector space in which the coordinates and velocities operate. The subspace defined by the eigenvectors of the constants of the motion must then be related to a Hilbert space, if our dynamical principle is to be internally consistent.

Remark on Cosmological Models

R. W. BASS AND L. WITTEN
RIAS, Inc., 2120 North Charles Street, Baltimore, Maryland

IT is unusual to hypothesize that the four-dimensional space-time universe of general relativity is compact (i.e., "finite"). But in such a case several interesting conclusions can be drawn. In the first place, if the mass distribution is assumed to be continuous, so that the metric tensor has no singularities, then the Euler-Poincaré characteristic of the universe must be zero.[1] This implies, for example, that *the universe cannot be a four-dimensional sphere*. It also implies that *a finite universe cannot be simply-connected*, in the sense that its first Betti number cannot vanish. (This is reminiscent of Professor J. A. Wheeler's nonsimply-connected models.)

In the second place, it seems to be generally known that *in a finite cosmology there must exist a closed curve in space-time whose tangent vector at every point is time-like*. Professor L. Markus has indicated a proof to us. Let V_4 denote the 4-manifold of the universe. Now, on V_4 construct a continuous, nowhere vanishing field of time-like vectors.[2,3] By Birkhoff's fundamental theorem on the existence of recurrent orbits in compact dynamical systems,[4] there must exist either an orbit of the type sought or else an "almost-closed" time-like orbit which can serve for the construction of such a closed orbit by an obvious procedure.

A more standard hypothesis, however, is that the universe V_4 is not compact, but is the topological product of the infinite real line (a time axis) with a 3-manifold V_3. The manifold V_3 is often assumed to be compact, and any local (hence experimentally verifiable) condition which implies compactness is of much interest. For example, if V_3 has constant curvature K, then V_3 is compact if K is positive,[5] and in this case is a 3-sphere if its first Betti number vanishes, and in general admits the 3-sphere as a covering space.

We wish to point out a new method for studying the topology of manifolds such as V_3 and V_4. This method consists of the construction of a continuous, nowhere vanishing, irrotational vector field on the manifold under consideration. Once such a vector field has been constructed, we can assert that either the manifold is noncompact (i.e., open or "infinite"), or that it cannot be simply connected.

We shall prove a slight generalization of this theorem; but first, let us note that a similar, but much more restrictive and less easily applicable condition is a trivial consequence of Hodge's well-known theorem that the number of linearly independent harmonic vector fields on a compact Riemannian manifold is equal to its first Betti number. For if after constructing on our manifold an irrotational vector field (which is nontrivial, but may vanish at more than one point), we then ascertain that it is also solenoidal (i.e., of vanishing divergence), then the vector field must be harmonic.[6]

Theorem 1 (Hodge). *Let V_n be an n-dimensional Riemannian manifold* (with positive definite metric tensor), *and let F denote a nontrivial class C_2 vector field defined on V_n. Suppose that the curl and the divergence of F both vanish identically; or equivalently, suppose that the field F satisfies the generalized Laplace equation for harmonic vector fields. Then, if V_n is compact, its first Betti number is not zero.*

Corollary (Bochner-Myers). *If V_n is orientable and has positive definite Ricci curvature throughout, then its first Betti number vanishes.*[7]

Recall that the curl tensor of a vector field is independent of the metric tensor, and so is a nonmetric notion. Accordingly, *the following theorem applies equally well to V_4 with its indefinite hyperbolic metric as to V_3 with its positive definite Riemannian metric.*

Theorem 2. *Let V_n be an n-dimensional differentiable manifold, and let F be a continuous, class C^1 vector field defined on V_n. Suppose that F vanishes at most once and that its curl vanishes identically on V_n. Then, either V_n is noncompact, or V_n is compact and its first Betti number does not vanish. In either case, of course, if F actually vanishes nowhere, the Euler-Poincaré characteristic of V_n is zero.*

For nonvanishing F this theorem is a consequence of a more general theorem[8] which applies, for example, to manifolds with boundary. In fact, by a generalization to arbitrary flows of a theorem proved by Lichnerowicz for a very special class of flows,[9] we can prove[10] that V_n is homeomorphic to the product of the real line with an $(n-1)$-dimensional space which is a connected subset of a V_{n-1}. But in the present case, because we are deal-

---

[1] L. Markus, Ann. Math. 62, 411–417 (1955).
[2] A. Lichnerowicz, *Théories Relativistes de la Gravitation et de l'Électromagnétism* (Masson et Cie, Paris, 1955), pp. 6, 7.
[3] N. Steenrod, *The Topology of Fibre Bundles* (Princeton University Press, Princeton, 1951), p. 207.
[4] G. D. Birkhoff, *Dynamical Systems* (American Mathematical Society, New York, 1927).
[5] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, 1949), pp. 84, 203.
[6] K. Yano and S. Bochner, "Curvature and Betti Numbers," Annals of Mathematics Studies No. 32, Princeton, 1953, p. 56.
[7] See reference 6, p. 37.
[8] R. W. Bass, "A criterion for the existence of unstable sets," a RIAS preprint, December 26, 1956. Copies available on request.
[9] See reference 2, p. 79.
[10] R. W. Bass, "Topological dynamics and nonlinear differential equations," a RIAS preprint (to be published).

ing with a manifold, there is a much simpler proof. We wish to thank Professor Kervaire for pointing out to us this simpler proof during the 1957 North Carolina Conference on Gravitation and General Relativity. The proof runs as follows. If V_n is simply connected, then the generalized Stokes theorem assures us that there exists on V_n a single-valued scalar potential function of which F is the gradient field. (See the survey of vector analysis in the paper by H. Weyl.[11]) But if V_n is compact, this potential function must assume both its maximum and minimum values on M, and at these extreme points the gradient must vanish. This contradicts the hypothesis that F has at most one zero on V_n, and so proves the theorem.

It is possible that Theorems 1 and 2 have applications to the study of specific cosmological models. In fact, there are many ways of constructing on V_3 or on V_4 continuous vector fields which are unique once the indefinite metric (or set of gravitational potentials) for V_4 has been specified.

Professor J. A. Wheeler has pointed out to us an application of Theorem 2 to V_4.

Theorem 3. *Consider the combined Einstein-Maxwell field theory on V_4. If the vector field $u_i = \epsilon_{ijkl} R_m{}^{j;l} R^{mk} / R_{pq} R^{pq}$ is defined every where and of class C^1 on V_4, and if u_i does not vanish more than once, then the universe V_4 cannot be compact.*

The vector field u_i, which was defined by Dr. C. Misner, is essentially the gradient of the ratio (in a certain coordinate system) of the electric to the magnetic field strength. Dr. Misner has shown that Maxwell's equations imply that $u_{i;j} - u_{j;i} \equiv 0$. Hence, Theorem 3 follows from Theorem 2.

[11] H. Weyl, Duke Math. J. 7, 411–444 (1940).

"Relative State" Formulation of Quantum Mechanics*

HUGH EVERETT, III†

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

1. INTRODUCTION

THE task of quantizing general relativity raises serious questions about the meaning of the present formulation and interpretation of quantum mechanics when applied to so fundamental a structure as the space-time geometry itself. This paper seeks to clarify the foundations of quantum mechanics. It presents a reformulation of quantum theory in a form believed suitable for application to general relativity.

The aim is not to deny or contradict the conventional formulation of quantum theory, which has demonstrated its usefulness in an overwhelming variety of problems, but rather to supply a new, more general and complete formulation, from which the conventional interpretation can be *deduced*.

The relationship of this new formulation to the older formulation is therefore that of a metatheory to a theory, that is, it is an underlying theory in which the nature and consistency, as well as the realm of applicability, of the older theory can be investigated and clarified.

The new theory is not based on any radical departure from the conventional one. The special postulates in the old theory which deal with observation are omitted in the new theory. The altered theory thereby acquires a new character. It has to be analyzed in and for itself before any identification becomes possible between the quantities of the theory and the properties of the world of experience. The identification, when made, leads back to the omitted postulates of the conventional theory that deal with observation, but in a manner which clarifies their role and logical position.

We begin with a brief discussion of the conventional formulation, and some of the reasons which motivate one to seek a modification.

2. REALM OF APPLICABILITY OF THE CONVENTIONAL OR "EXTERNAL OBSERVATION" FORMULATION OF QUANTUM MECHANICS

We take the conventional or "external observation" formulation of quantum mechanics to be essentially the following[1]: A physical system is completely described by a state function ψ, which is an element of a Hilbert space, and which furthermore gives information only to the extent of specifying the probabilities of the results of various observations which can be made *on* the system *by* external observers. There are two fundamentally different ways in which the state function can change:

Process 1: The discontinuous change brought about by the observation of a quantity with eigenstates $\phi_1, \phi_2, \cdots$, in which the state ψ will be changed to the state ϕ_j with probability $|(\psi,\phi_j)|^2$.

Process 2: The continuous, deterministic change of state of an isolated system with time according to a wave equation $\partial\psi/\partial t = A\psi$, where A is a linear operator.

This formulation describes a wealth of experience. No experimental evidence is known which contradicts it.

Not all conceivable situations fit the framework of this mathematical formulation. Consider for example an isolated system consisting of an observer or measuring apparatus, plus an object system. Can the change with time of the state of the *total* system be described by Process 2? If so, then it would appear that no discontinuous probabilistic process like Process 1 can take place. If not, we are forced to admit that systems which contain observers are not subject to the same kind of quantum-mechanical description as we admit for all other physical systems. The question cannot be ruled out as lying in the domain of psychology. Much of the discussion of "observers" in quantum mechanics has to do with photoelectric cells, photographic plates, and similar devices where a mechanistic attitude can hardly be contested. For the following one can *limit himself to this class of problems*, if he is unwilling to consider observers in the more familiar sense on the same mechanistic level of analysis.

What mixture of Processes 1 and 2 of the conventional formulation is to be applied to the case where only an approximate measurement is effected; that is, where an apparatus or observer interacts only weakly and for a limited time with an object system? In this case of an

* Thesis submitted to Princeton University March 1, 1957 in partial fulfillment of the requirements for the Ph.D. degree. An earlier draft dated January, 1956 was circulated to several physicists whose comments were helpful. Professor Niels Bohr, Dr. H. J. Groenewald, Dr. Aage Peterson, Dr. A. Stern, and Professor L. Rosenfeld are free of any responsibility, but they are warmly thanked for the useful objections that they raised. Most particular thanks are due to Professor John A. Wheeler for his continued guidance and encouragement. Appreciation is also expressed to the National Science Foundation for fellowship support.

† Present address: Weapons Systems Evaluation Group, The Pentagon, Washington, D. C.

[1] We use the terminology and notation of J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton University Press, Princeton, 1955).

approximate measurement a proper theory must specify (1) the new state of the object system that corresponds to any particular reading of the apparatus and (2) the probability with which this reading will occur. von Neumann showed how to treat a special class of approximate measurements by the method of projection operators.[2] However, a general treatment of all approximate measurements by the method of projection operators can be shown (Sec. 4) to be impossible.

How is one to apply the conventional formulation of quantum mechanics to the space-time geometry itself? The issue becomes especially acute in the case of a closed universe.[3] There is no place to stand outside the system to observe it. There is nothing outside it to produce transitions from one state to another. Even the familiar concept of a proper state of the energy is completely inapplicable. In the derivation of the law of conservation of energy, one defines the total energy by way of an integral extended over a surface large enough to include all parts of the system and their interactions.[4] But in a closed space, when a surface is made to include more and more of the volume, it ultimately disappears into nothingness. Attempts to define a total energy for a closed space collapse to the vacuous statement, zero equals zero.

How are a quantum description of a closed universe, of approximate measurements, and of a system that contains an observer to be made? These three questions have one feature in common, that they all inquire about the *quantum mechanics* that is *internal to an isolated system*.

No way is evident to apply the conventional formulation of quantum mechanics to a system that is not subject to *external* observation. The whole interpretive scheme of that formalism rests upon the notion of external observation. The probabilities of the various possible outcomes of the observation are prescribed exclusively by Process 1. Without that part of the formalism there is no means whatever to ascribe a physical interpretation to the conventional machinery. But Process 1 is out of the question for systems not subject to external observation.[5]

3. QUANTUM MECHANICS INTERNAL TO AN ISOLATED SYSTEM

This paper proposes to regard pure wave mechanics (Process 2 only) as a complete theory. It postulates that a wave function that obeys a linear wave equation everywhere and at all times supplies a complete mathematical model for every isolated physical system without exception. It further postulates that every system that is subject to external observation can be regarded as part of a larger isolated system.

The wave function is taken as the basic physical entity with *no a priori interpretation*. Interpretation only comes *after* an investigation of the logical structure of the theory. Here as always the theory itself sets the framework for its interpretation.[5]

For any interpretation it is necessary to put the mathematical model of the theory into correspondence with experience. For this purpose it is necessary to formulate abstract models for observers that can be treated within the theory itself as physical systems, to consider isolated systems containing such model observers in interaction with other subsystems, to deduce the changes that occur in an observer as a consequence of interaction with the surrounding subsystems, and to interpret the changes in the familiar language of experience.

Section 4 investigates representations of the state of a composite system in terms of states of constituent subsystems. The mathematics leads one to recognize the concept of the *relativity of states*, in the following sense: a constituent subsystem cannot be said to be in any single well-defined state, independently of the remainder of the composite system. To any arbitrarily chosen state for one subsystem there will correspond a unique *relative state* for the remainder of the composite system. This relative state will usually depend upon the choice of state for the first subsystem. Thus the state of one subsystem does not have an independent existence, but is fixed only by the state of the remaining subsystem. In other words, the states occupied by the subsystems are not independent, but *correlated*. Such correlations between systems arise whenever systems interact. In the present formulation all measurements and observation processes are to be regarded simply as interactions between the physical systems involved—interactions which produce strong correlations. A simple model for a measurement, due to von Neumann, is analyzed from this viewpoint.

Section 5 gives an abstract treatment of the problem of observation. This uses only the superposition principle, and general rules by which composite system states are formed of subsystem states, in order that the results shall have the greatest generality and be applicable to any form of quantum theory for which these principles hold. Deductions are drawn about the state of the observer relative to the state of the object system. It is found that experiences of the observer (magnetic tape memory, counter system, etc.) are in full accord with predictions of the conventional "external observer" formulation of quantum mechanics, based on Process 1.

Section 6 recapitulates the "relative state" formulation of quantum mechanics.

[2] Reference 1, Chap. 4, Sec. 4.
[3] See A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, 1950), third edition, p. 107.
[4] L. Landau and E. Lifshitz, *The Classical Theory of Fields*, translated by M. Hamermesh (Addison-Wesley Press, Cambridge, 1951), p. 343.
[5] See in particular the discussion of this point by N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **12**, No. 8 (1933).

4. CONCEPT OF RELATIVE STATE

We now investigate some consequences of the wave mechanical formalism of composite systems. If a composite system S, is composed of two subsystems S_1 and S_2, with associated Hilbert spaces H_1 and H_2, then, according to the usual formalism of composite systems, the Hilbert space for S is taken to be the *tensor product* of H_1 and H_2 (written $H=H_1\otimes H_2$). This has the consequence that if the sets $\{\xi_i^{S_1}\}$ and $\{\eta_j^{S_2}\}$ are complete orthonormal sets of states for S_1 and S_2, respectively, then the general state of S can be written as a superposition:

$$\psi^S=\sum_{i,j}a_{ij}\xi_i^{S_1}\eta_j^{S_2}. \qquad (1)$$

From (3.1) although S is in a definite state ψ^S, the subsystems S_1 and S_2 do not possess anything like definite states independently of one another (except in the special case where all but one of the a_{ij} are zero).

We *can*, however, for any choice of a state in one subsystem, *uniquely* assign a corresponding *relative* state in the other subsystem. For example, if we choose ξ_k as the state for S_1, while the composite system S is in the state ψ^S given by (3.1), then the corresponding *relative state* in S_2, $\psi(S_2;\text{rel}\xi_k,S_1)$, will be:

$$\psi(S_2;\text{rel}\xi_k,S_1)=N_k\sum_j a_{kj}\eta_j^{S_2} \qquad (2)$$

where N_k is a normalization constant. This relative state for ξ_k is *independent* of the choice of basis $\{\xi_i\}$ $(i\neq k)$ for the orthogonal complement of ξ_k, and is hence determined uniquely by ξ_k alone. To find the relative state in S_2 for an arbitrary state of S_1 therefore, one simply carries out the above procedure using any pair of bases for S_1 and S_2 which contains the desired state as one element of the basis for S_1. To find states in S_1 relative to states in S_2, interchange S_1 and S_2 in the procedure.

In the conventional or "external observation" formulation, the relative state in S_2, $\psi(S_2;\text{rel}\phi,S_1)$, for a state ϕ^{S_1} in S_1, gives the conditional probability distributions for the results of all measurements in S_2, given that S_1 has been measured and found to be in state ϕ^{S_1}—i.e., that ϕ^{S_1} is the eigenfunction of the measurement in S_1 corresponding to the observed eigenvalue.

For any choice of basis in S_1, $\{\xi_i\}$, it is always possible to represent the state of S, (1), as a *single* superposition of pairs of states, each consisting of a state from the basis $\{\xi_i\}$ in S_1 and its relative state in S_2. Thus, from (2), (1) can be written in the form:

$$\psi^S=\sum_i\frac{1}{N_i}\xi_i^{S_1}\psi(S_2;\text{rel}\xi_i,S_1). \qquad (3)$$

This is an important representation used frequently.

Summarizing: There does not, in general, exist anything like a single state for one subsystem of a composite system. Subsystems do not possess states that are independent of the states of the remainder of the system, so that the subsystem states are generally correlated *with one another. One can arbitrarily choose a state for one subsystem, and be led to the relative state for the remainder. Thus we are faced with a fundamental* relativity of states, *which is implied by the formalism of composite systems. It is meaningless to ask the absolute state of a subsystem—one can only ask the state relative to a given state of the remainder of the subsystem.*

At this point we consider a simple example, due to von Neumann, which serves as a model of a measurement process. Discussion of this example prepares the ground for the analysis of "observation." We start with a system of only one coordinate, q (such as position of a particle), and an apparatus of one coordinate r (for example the position of a meter needle). Further suppose that they are initially independent, so that the combined wave function is $\psi_0^{S+A}=\phi(q)\eta(r)$ where $\phi(q)$ is the initial system wave function, and $\eta(r)$ is the initial apparatus function. The Hamiltonian is such that the two systems do not interact except during the interval $t=0$ to $t=T$, during which time the total Hamiltonian consists only of a simple interaction,

$$H_I=-i\hbar q(\partial/\partial r). \qquad (4)$$

Then the state

$$\psi_t^{S+A}(q,r)=\phi(q)\eta(r-qt) \qquad (5)$$

is a solution of the Schrödinger equation,

$$i\hbar(\partial\psi_t^{S+A}/\partial t)=H_I\psi_t^{S+A}, \qquad (6)$$

for the specified initial conditions at time $t=0$.

From (5) at time $t=T$ (at which time interaction stops) there is no longer any definite independent apparatus state, nor any independent system state. The apparatus therefore does not indicate any definite object-system value, and nothing like process 1 has occurred.

Nevertheless, we *can* look upon the total wave function (5) as a *superposition* of pairs of subsystem states, each element of which has a definite q value and a correspondingly displaced apparatus state. Thus after the interaction the state (5) has the form:

$$\psi_T^{S+A}=\int\phi(q')\delta(q-q')\eta(r-qT)dq', \qquad (7)$$

which is a superposition of states $\psi_{q'}=\delta(q-q')\eta(r-qT)$. Each of these elements, $\psi_{q'}$, of the superposition describes a state in which the system has the definite value $q=q'$, and in which the apparatus has a state that is displaced from its original state by the amount $q'T$. These elements $\psi_{q'}$ are then superposed with coefficients $\phi(q')$ to form the total state (7).

Conversely, if we transform to the representation where the *apparatus* coordinate is definite, we write (5) as

$$\psi_T^{S+A}=\int(1/N_{r'})\xi^{r'}(q)\delta(r-r')dr',$$

where

$$\xi^{r'}(q) = N_{r'}\phi(q)\eta(r'-qT) \tag{8}$$

and

$$(1/N_{r'})^2 = \int \phi^*(q)\phi(q)\eta^*(r'-qT)\eta(r'-qT)dq.$$

Then the $\xi^{r'}(q)$ are the relative system state functions[6] for the apparatus states $\delta(r-r')$ of definite value $r=r'$.

If T is sufficiently large, or $\eta(r)$ sufficiently sharp (near $\delta(r)$), then $\xi^{r'}(q)$ is nearly $\delta(q-r'/T)$ and the relative system states $\xi^{r'}(q)$ are nearly eigenstates for the values $q=r'/T$.

We have seen that (8) is a superposition of states $\psi_{r'}$, *for each of which* the apparatus has recorded a definite value r', and the system is left in approximately the eigenstate of the measurement corresponding to $q=r'/T$. The discontinuous "jump" into an eigenstate is thus only a relative proposition, dependent upon the mode of decomposition of the total wave function into the superposition, and relative to a particularly chosen apparatus-coordinate value. So far as the complete theory is concerned all elements of the superposition exist simultaneously, and the entire process is quite continuous.

von Neumann's example is only a special case of a more general situation. Consider any measuring apparatus interacting with any object system. As a result of the interaction the state of the measuring apparatus is no longer capable of independent definition. It can be defined only *relative* to the state of the object system. In other words, there exists only a correlation between the states of the two systems. It seems as if nothing can ever be settled by such a measurement.

This indefinite behavior seems to be quite at variance with our observations, since physical objects always appear to us to have definite positions. Can we reconcile this feature wave mechanical theory built purely on Process 2 with experience, or must the theory be abandoned as untenable? In order to answer this question we consider the problem of observation itself within the framework of the theory.

5. OBSERVATION

We have the task of making deductions about the appearance of phenomena to observers which are considered as purely physical systems and are treated within the theory. To accomplish this it is necessary to identify some present properties of such an observer with features of the past experience of the observer. Thus, in order to say that an observer 0 has observed the event α, it is necessary that the state of 0 has become changed from its former state to a new state which is dependent upon α.

It will suffice for our purposes to consider the observers to possess memories (i.e., parts of a relatively permanent nature whose states are in correspondence with past experience of the observers). In order to make deductions about the past experience of an observer it is sufficient to deduce the present contents of the memory as it appears within the mathematical model.

As models for observers we can, if we wish, consider automatically functioning machines, possessing sensory apparatus and coupled to recording devices capable of registering past sensory data and machine configurations. We can further suppose that the machine is so constructed that its present actions shall be determined not only by its present sensory data, but by the contents of its memory as well. Such a machine will then be capable of performing a sequence of observations (measurements), and furthermore of deciding upon its future experiments on the basis of past results. If we consider that current sensory data, as well as machine configuration, is immediately recorded in the memory, then the actions of the machine at a given instant can be regarded as a function of the memory contents only, and all relavant experience of the machine is contained in the memory.

For such machines we are justified in using such phrases as "the machine has perceived A" or "the machine is aware of A" if the occurrence of A is represented in the memory, since the future behavior of the machine will be based upon the occurrence of A. In fact, all of the customary language of subjective experience is quite applicable to such machines, and forms the most natural and useful mode of expression when dealing with their behavior, as is well known to individuals who work with complex automata.

When dealing with a system representing an observer quantum mechanically we ascribe a state function, ψ^0, to it. When the state ψ^0 describes an observer whose memory contains representations of the events $A, B, \cdots, C$ we denote this fact by appending the memory sequence in brackets as a subscript, writing:

$$\psi^0{}_{[A,B,\cdots,C]}. \tag{9}$$

The symbols $A, B, \cdots, C$, which we assume to be ordered time-wise, therefore stand for memory configurations which are in correspondence with the past experience of the observer. These configurations can be regarded as punches in a paper tape, impressions on a magnetic reel, configurations of a relay switching circuit, or even configurations of brain cells. We require only that they be capable of the interpretation "The observer has experienced the succession of events $A, B, \cdots, C$." (We sometimes write dots in a memory sequence, $\cdots A, B, \cdots C$, to indicate the possible presence of previous

[6] This example provides a model of an approximate measurement. However, the relative system states after the interaction $\xi^{r'}(q)$ cannot ordinarily be generated from the original system state ϕ by the application of *any* projection operator, E. Proof: Suppose on the contrary that $\xi^{r'}(q) = NE\phi(q) = N'\phi(q)\eta(r'-qt)$, where N, N' are normalization constants. Then

$$E(NE\phi(q)) = NE^2\phi(q) = N''\phi(q)\eta^2(r'-qt)$$

and $E^2\phi(q) = (N''/N)\phi(q)\eta^2(r'-qt)$. But the condition $E^2=E$ which is necessary for E to be a projection implies that N'/N'' $\eta(q)=\eta^2(q)$ which is generally false.

memories which are irrelevant to the case being considered.)

The mathematical model seeks to treat the interaction of such observer systems with other physical systems (observations), within the framework of Process 2 wave mechanics, and to deduce the resulting memory configurations, which are then to be interpreted as records of the past experiences of the observers.

We begin by defining what constitutes a "good" observation. A good observation of a quantity A, with eigenfunctions ϕ_i, for a system S, by an observer whose initial state is ψ^0, consists of an interaction which, in a specified period of time, transforms each (total) state

$$\psi^{S+0}=\phi_i\psi^0[\ldots] \tag{10}$$

into a new state

$$\psi^{S+0'}=\phi_i\psi^0[\ldots\alpha_i] \tag{11}$$

where α_i characterizes[7] the state ϕ_i. (The symbol, α_i, might stand for a recording of the eigenvalue, for example.) That is, we require that the system state, *if it is an eigenstate*, shall be unchanged, and (2) that the observer state shall change so as to describe an observer that is "aware" of which eigenfunction it is; that is, some property is recorded in the memory of the observer which characterizes ϕ_i, such as the eigenvalue. The requirement that the eigenstates for the system be unchanged is necessary if the observation is to be significant (repeatable), and the requirement that the observer state change in a manner which is different for each eigenfunction is necessary if we are to be able to call the interaction an observation at all. How closely a general interaction satisfies the definition of a good observation depends upon (1) the way in which the interaction depends upon the dynamical variables of the observer system—including memory variables—and upon the dynamical variables of the object system and (2) the initial state of the observer system. Given (1) and (2), one can for example solve the wave equation, deduce the state of the composite system after the end of the interaction, and check whether an object system that was originally in an eigenstate is left in an eigenstate, as demanded by the repeatability postulate. This postulate is satisfied, for example, by the model of von Neumann that has already been discussed.

From the definition of a good observation we first deduce the result of an observation upon a system which is *not* in an eigenstate of the observation. We know from our definition that the interaction transforms states $\phi_i\psi^0[\cdots]$ into states $\phi_i\psi^0[\cdots\alpha_i]$. Consequently these solutions of the wave equation can be superposed to give the final state for the case of an arbitrary initial system state. Thus if the initial system state is not an eigenstate, but a general state $\sum_i a_i\phi_i$, the final total state will have the form:

$$\psi^{S+0'}=\sum a_i\phi_i\psi^0[\cdots\alpha_i]. \tag{12}$$

This superposition principle continues to apply in the presence of further systems which do not interact during the measurement. Thus, if systems $S_1, S_2, \cdots, S_n$ are present as well as 0, with original states $\psi^{S_1},\psi^{S_2}, \cdots,\psi^{S_n}$, and the only interaction during the time of measurement takes place between S_1 and 0, the measurement will transform the initial total state:

$$\psi^{S_1+S_2+\cdots+S_n+0}=\psi^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0[\cdots] \tag{13}$$

into the final state:

$$\psi'^{S_1+S_2+\cdots+S_n+0}=\sum_i a_i\phi_i^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0[\cdots\alpha_i] \tag{14}$$

where $a_i=(\phi_i^{S_1},\psi^{S_1})$ and $\phi_i^{S_1}$ are eigenfunctions of the observation.

Thus we arrive at the general rule for the transformation of total state functions which describe systems within which observation processes occur:

Rule 1: The observation of a quantity A, with eigenfunctions $\phi_i^{S_1}$, in a system S_1 by the observer 0, transforms the total state according to:

$$\psi^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0[\cdots] \rightarrow\sum_i a_i\phi_i^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0[\cdots\alpha_i] \tag{15}$$

where

$$a_i=(\phi_i^{S_1},\psi^{S_1}).$$

If we next consider a *second* observation to be made, where our total state is now a superposition, we can apply Rule 1 separately to each element of the superposition, since each element separately obeys the wave equation and behaves independently of the remaining elements, and then superpose the results to obtain the final solution. We formulate this as:

Rule 2: Rule 1 may be applied separately to each element of a superposition of total system states, the results being superposed to obtain the final total state. Thus, a determination of B, with eigenfunctions $\eta_j^{S_2}$, on S_2 by the observer 0 transforms the total state

$$\sum_i a_i\phi_i^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0[\cdots\alpha_i] \tag{16}$$

into the state

$$\sum_{i,j} a_i b_j\phi_i^{S_1}\eta_j^{S_2}\psi^{S_3}\cdots\psi^{S_n}\psi^0[\cdots\alpha_i,\beta_j] \tag{17}$$

where $b_j=(\eta_j^{S_2},\psi^{S_2})$, which follows from the application of Rule 1 to each element $\phi_i^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0[\cdots\alpha_i]$, and then superposing the results with the coefficients a_i.

These two rules, which follow directly from the superposition principle, give a convenient method for determining final total states for any number of observation processes in any combinations. We now seek the *interpretation* of such final total states.

[7] It should be understood that $\psi^0[\cdots\alpha_i]$ is a *different* state for each i. A more precise notation would write $\psi^0_i[\cdots\alpha_i]$, but no confusion can arise if we simply let the ψ_i^0 be indexed only by the index of the memory configuration symbol.

Let us consider the simple case of a single observation of a quantity A, with eigenfunctions ϕ_i, in the system S with initial state ψ^S, by an observer 0 whose initial state is $\psi^0{}_{[\cdots]}$. The final result is, as we have seen, the superposition

$$\psi'^{S+0}=\sum_i a_i\phi_i\psi^0{}_{[\cdots\alpha_i]}. \quad (18)$$

There is no longer any independent system state or observer state, although the two have become correlated in a one-one manner. However, in each *element* of the superposition, $\phi_i\psi^0{}_{[\cdots\alpha_i]}$, the object-system state is a particular eigenstate of the observation, and *furthermore the observer-system state describes the observer as definitely perceiving that particular system state.* This correlation is what allows one to maintain the interpretation that a measurement has been performed.

We now consider a situation where the observer system comes into interaction with the object system for a second time. According to Rule 2 we arrive at the total state after the second observation:

$$\psi''^{S+0}=\sum_i a_i\phi_i\psi^0{}_{[\cdots\alpha_i,\alpha_i]}. \quad (19)$$

Again, each element $\phi_i\psi^0{}_{[\cdots\alpha_i,\alpha_i]}$ describes a system eigenstate, but this time also describes the observer as having obtained the *same result* for each of the two observations. Thus for every separate state of the observer in the final superposition the result of the observation was repeatable, even though different for different states. This repeatability is a consequence of the fact that after an observation the *relative* system state for a particular observer state is the corresponding eigenstate.

Consider now a different situation. An observer-system 0, with initial state $\psi^0{}_{[\cdots]}$, measures the *same* quantity A in a number of separate, identical, systems which are initially in the same state, $\psi^{S_1}=\psi^{S_2}=\cdots=\psi^{S_n}=\sum_i a_i\phi_i$ (where the ϕ_i are, as usual, eigenfunctions of A). The initial total state function is then

$$\psi_0{}^{S_1+S_2+\cdots+S_n+0}=\psi^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0{}_{[\cdots]}. \quad (20)$$

We assume that the measurements are performed on the systems in the order $S_1, S_2, \cdots S_n$. Then the total state after the first measurement is by Rule 1,

$$\psi_1{}^{S_1+S_2+\cdots+S_n+0}=\sum_i a_i\phi_i{}^{S_1}\psi^{S_2}\cdots\psi^{S_n}\psi^0{}_{[\cdots\alpha_i{}^1]} \quad (21)$$

(where $\alpha_i{}^1$ refers to the first system, S_1).

After the second measurement it is, by Rule 2,

$$\psi_2{}^{S_1+S_2+\cdots+S_n+0}=\sum_{i,j}a_ia_j\phi_i{}^{S_1}\phi_j{}^{S_2}\psi^{S_3}\cdots\psi^{S_n}\psi^0{}_{[\cdots\alpha_i{}^1,\alpha_j{}^2]} \quad (22)$$

and in general, after r measurements have taken place ($r\leq n$), Rule 2 gives the result:

$$\psi_r=\sum_{i,j,\ldots k}a_ia_j\cdots a_k\phi_i{}^{S_1}\phi_j{}^{S_2}\cdots\phi_k{}^{S_r}\psi^{S_{r+1}}\cdots\psi^{S_n}\psi^0{}_{[\cdots\alpha_i{}^1,\alpha_j{}^2,\cdots\alpha_k{}^r]} \quad (23)$$

We can give this state, ψ_r, the following interpretation. It consists of a superposition of states:

$$\psi'_{ij\ldots k}=\phi_i{}^{S_1}\phi_j{}^{S_2}\cdots\phi_k{}^{S_r}\times\psi^{S_{r+1}}\cdots\psi^{S_n}\psi^0{}_{[\alpha_i{}^1,\alpha_j{}^2,\cdots\alpha_k{}^r]} \quad (24)$$

each of which describes the observer with a definite memory sequence $[\alpha_i{}^1,\alpha_j{}^2,\cdots\alpha_k{}^r]$. Relative to him the (observed) system states are the corresponding eigenfunctions $\phi_i{}^{S_1}, \phi_j{}^{S_2},\cdots, \phi_k{}^{S_r}$, the remaining systems, $S_{r+1},\cdots, S_n$, being unaltered.

A typical element $\psi'_{ij\ldots k}$ of the final superposition describes a state of affairs wherein the observer has perceived an apparently random sequence of definite results for the observations. Furthermore the object systems have been left in the corresponding eigenstates of the observation. At this stage suppose that a redetermination of an earlier system observation (S_l) takes place. Then it follows that every element of the resulting final superposition will describe the observer with a memory configuration of the form $[\alpha_i{}^1,\cdots\alpha_j{}^l,\cdots\alpha_k{}^r,\alpha_j{}^l]$ in which the earlier memory coincides with the later—i.e., the memory states are *correlated.* It will thus *appear* to the observer, as described by a typical element of the superposition, that each initial observation on a system caused the system to "jump" into an eigenstate in a random fashion and thereafter remain there for subsequent measurements on the same system. Therefore—disregarding for the moment quantitative questions of relative frequencies—the probabilistic assertions of Process 1 *appear* to be valid to the observer described by a typical element of the final superposition.

We thus arrive at the following picture: Throughout all of a sequence of observation processes there is only one physical system representing the observer, yet there is no single unique *state* of the observer (which follows from the representations of interacting systems). Nevertheless, there is a representation in terms of a *superposition*, each element of which contains a definite observer state and a corresponding system state. Thus with each succeeding observation (or interaction), the observer state "branches" into a number of different states. Each branch represents a different outcome of the measurement and the *corresponding* eigenstate for the object-system state. All branches exist simultaneously in the superposition after any given sequence of observations.‡

‡ *Note added in proof.*—In reply to a preprint of this article some correspondents have raised the question of the "transition from possible to actual," arguing that in "reality" there is—as our experience testifies—no such splitting of observer states, so that only one branch can ever actually exist. Since this point may occur to other readers the following is offered in explanation.

The whole issue of the transition from "possible" to "actual" is taken care of in the theory in a very simple way—there is no such transition, nor is such a transition necessary for the theory to be in accord with our experience. From the viewpoint of the theory *all* elements of a superposition (all "branches") are "actual," none any more "real" than the rest. It is unnecessary to suppose that all but one are somehow destroyed, since all the

The "trajectory" of the memory configuration of an observer performing a sequence of measurements is thus not a linear sequence of memory configurations, but a branching tree, with all possible outcomes existing simultaneously in a final superposition with various coefficients in the mathematical model. In any familiar memory device the branching does not continue indefinitely, but must stop at a point limited by the capacity of the memory.

In order to establish quantitative results, we must put some sort of measure (weighting) on the elements of a final superposition. This is necessary to be able to make assertions which hold for almost all of the observer states described by elements of a superposition. We wish to make quantitative statements about the relative frequencies of the different possible results of observation—which are recorded in the memory—for a typical observer state; but to accomplish this we must have a method for selecting a typical element from a superposition of orthogonal states.

We therefore seek a general scheme to assign a measure to the elements of a superposition of orthogonal states $\sum_i a_i\phi_i$. We require a positive function m of the complex coefficients of the elements of the superposition, so that $m(a_i)$ shall be the measure assigned to the element ϕ_i. In order that this general scheme be unambiguous we must first require that the states themselves always be normalized, so that we can distinguish the coefficients from the states. However, we can still only determine the *coefficients*, in distinction to the states, up to an arbitrary phase factor. In order to avoid ambiguities the function m must therefore be a function of the amplitudes of the coefficients alone, $m(a_i)=m(|a_i|)$.

We now impose an additivity requirement. We can regard a subset of the superposition, say $\sum_{i=1}^{n} a_i\phi_i$, as a single element $\alpha\phi'$:

$$\alpha\phi' = \sum_{i=1}^{n} a_i\phi_i. \tag{25}$$

We then demand that the measure assigned to ϕ' shall be the sum of the measures assigned to the ϕ_i (i from 1 to n):

$$m(\alpha) = \sum_{i=1}^{n} m(a_i). \tag{26}$$

Then we have already restricted the choice of m to the square amplitude alone; in other words, we have $m(a_i)=a_i^*a_i$, apart from a multiplicative constant.

To see this, note that the normality of ϕ' requires that $|\alpha|=(\sum a_i^*a_i)^{\frac{1}{2}}$. From our remarks about the dependence of m upon the amplitude alone, we replace the a_i by their amplitudes $u_i=|a_i|$. Equation (26) then imposes the requirement,

$$m(\alpha)=m(\textstyle\sum a_i^*a_i)^{\frac{1}{2}}=m(\sum u_i^2)^{\frac{1}{2}} \\ =\textstyle\sum m(u_i)=\sum m(u_i^2)^{\frac{1}{2}}. \tag{27}$$

Defining a new function $g(x)$

$$g(x)=m(\sqrt{x}) \tag{28}$$

we see that (27) requires that

$$g(\textstyle\sum u_i^2)=\sum g(u_i^2). \tag{29}$$

Thus g is restricted to be linear and necessarily has the form:

$$g(x)=cx \quad (c \text{ constant}). \tag{30}$$

Therefore $g(x^2)=cx^2=m(\sqrt{x^2})=m(x)$ and we have deduced that m is restricted to the form

$$m(a_i)=m(u_i)=cu_i^2=ca_i^*a_i. \tag{31}$$

We have thus shown that the only choice of measure consistent with our additivity requirement is the square amplitude measure, apart from an arbitrary multiplicative constant which may be fixed, if desired, by normalization requirements. (The requirement that the total measure be unity implies that this constant is 1.)

The situation here is fully analogous to that of classical statistical mechanics, where one puts a measure on trajectories of systems in the phase space by placing a measure on the phase space itself, and then making assertions (such as ergodicity, quasi-ergodicity, etc.) which hold for "almost all" trajectories. This notion of "almost all" depends here also upon the choice of measure, which is in this case taken to be the Lebesgue measure on the phase space. One could contradict the statements of classical statistical mechanics by choosing a measure for which only the exceptional trajectories had nonzero measure. Nevertheless the choice of Lebesgue measure on the phase space can be justified by the fact that it is the only choice for which the "conservation of probability" holds, (Liouville's theorem) and hence the only choice which makes possible any reasonable statistical deductions at all.

In our case, we wish to make statements about "trajectories" of observers. However, for us a trajectory is constantly branching (transforming from state to superposition) with each successive measurement. To

---

separate elements of a superposition individually obey the wave equation with complete indifference to the presence or absence ("actuality" or not) of any other elements. This total lack of effect of one branch on another also implies that no observer will ever be aware of any "splitting" process.

Arguments that the world picture presented by this theory is contradicted by experience, because we are unaware of any branching process, are like the criticism of the Copernican theory that the mobility of the earth as a real physical fact is incompatible with the common sense interpretation of nature because we feel no such motion. In both cases the argument fails when it is shown that the theory itself predicts that our experience will be what it in fact is. (In the Copernican case the addition of Newtonian physics was required to be able to show that the earth's inhabitants would be unaware of any motion of the earth.)

have a requirement analogous to the "conservation of probability" in the classical case, we demand that the measure assigned to a trajectory at one time shall equal the sum of the measures of its separate branches at a later time. This is precisely the additivity requirement which we imposed and which leads uniquely to the choice of square-amplitude measure. Our procedure is therefore quite as justified as that of classical statistical mechanics.

Having deduced that there is a unique measure which will satisfy our requirements, the square-amplitude measure, we continue our deduction. This measure then assigns to the $i, j, \cdots k$th element of the superposition (24),

$$\phi_i^{S_1}\phi_j^{S_2}\cdots\phi_k^{S_r}\psi^{S_{r+1}}\cdots\psi^{S_n}\psi^0{}_{[\alpha_i^1,\alpha_j^2,\cdots\alpha_k^r]} \tag{32}$$

the measure (weight)

$$M_{ij\ldots k}=(a_ia_j\cdots a_k)^*(a_ia_j\cdots a_k) \tag{33}$$

so that the observer state with memory configuration $[\alpha_i^1, \alpha_j^2, \cdots, \alpha_k^r]$ is assigned the measure $a_i^*a_ia_j^*a_j\cdots a_k^*a_k=M_{ij\ldots k}$. We see immediately that this is a product measure, namely,

$$M_{ij\ldots k}=M_iM_j\cdots M_k \tag{34}$$

where

$$M_l=a_l^*a_l$$

so that the measure assigned to a particular memory sequence $[\alpha_i^1, \alpha_j^2, \cdots, \alpha_k^r]$ is simply the product of the measures for the individual components of the memory sequence.

There is a direct correspondence of our measure structure to the probability theory of random sequences. *If we regard* the $M_{ij\ldots k}$ as probabilities for the sequences then the sequences are equivalent to the random sequences which are generated by ascribing to each term the *independent* probabilities $M_l=a_l^*a_l$. Now probability theory is equivalent to measure theory mathematically, so that we can make use of it, while keeping in mind that all results should be translated back to measure theoretic language.

Thus, in particular, if we consider the sequences to become longer and longer (more and more observations performed) *each* memory sequence of the final superposition will satisfy any given criterion for a randomly generated sequence, generated by the independent probabilities $a_l^*a_l$, except for a set of total measure which tends toward zero as the number of observations becomes unlimited. Hence all averages of functions over *any* memory sequence, including the special case of frequencies, can be computed from the probabilities $a_i^*a_i$, except for a set of memory sequences of measure zero. We have therefore shown that the statistical assertions of Process 1 will appear to be valid to the observer, *in almost all* elements of the superposition (24), in the limit as the number of observations goes to infinity.

While we have so far considered only sequences of observations of the same quantity upon identical systems, the result is equally true for arbitrary sequences of observations, as may be verified by writing more general sequences of measurements, and applying Rules 1 and 2 in the same manner as presented here.

We can therefore summarize the situation when the sequence of observations is arbitrary, when these observations are made upon the same or different systems in any order, and when the number of observations of each quantity in each system is very large, with the following result:

> Except for a set of memory sequences of measure nearly zero, the averages of any functions over a memory sequence can be calculated approximately by the use of the independent probabilities given by Process 1 for each initial observation, on a system, and by the use of the usual transition probabilities for succeeding observations upon the same system. In the limit, as the number of all types of observations goes to infinity the calculation is exact, and the exceptional set has measure zero.

This prescription for the calculation of averages over memory sequences by probabilities assigned to individual elements is precisely that of the conventional "external observation" theory (Process 1). Moreover, these predictions hold for almost all memory sequences. Therefore all predictions of the usual theory will appear to be valid to the observer in amost all observer states.

In particular, the uncertainty principle is never violated since the latest measurement upon a system supplies all possible information about the relative system state, so that there is no direct correlation between any earlier results of observation on the system, and the succeeding observation. Any observation of a quantity B, between two successive observations of quantity A (all on the same system) will destroy the one-one correspondence between the earlier and later memory states for the result of A. Thus for alternating observations of different quantities there are fundamental limitations upon the correlations between memory states for the same observed quantity, these limitations expressing the content of the uncertainty principle.

As a final step one may investigate the consequences of allowing several observer systems to interact with (observe) the same object system, as well as to interact with one another (communicate). The latter interaction can be treated simply as an interaction which correlates parts of the memory configuration of one observer with another. When these observer systems are investigated, in the same manner as we have already presented in this section using Rules 1 and 2, one finds that in *all elements* of the final superposition:

1. When several observers have separately observed the same quantity in the object system and then communicated the results to one another they find that they

are in agreement. This agreement persists even when an observer performs his observation *after* the result has been communicated to him by another observer who has performed the observation.

2. Let one observer perform an observation of a quantity A in the object system, then let a second perform an observation of a quantity B in this object system which does not commute with A, and finally let the first observer repeat his observation of A. Then the memory system of the first observer will *not* in general show the same result for both observations. The intervening observation by the other observer of the noncommuting quantity B prevents the possibility of any one to one correlation between the two observations of A.

3. Consider the case where the states of two object systems are correlated, but where the two systems do not interact. Let one observer perform a specified observation on the first system, then let another observer perform an observation on the second system, and finally let the first observer repeat his observation. Then it is found that the first observer always gets the same result both times, and the observation by the second observer has no effect whatsoever on the outcome of the first's observations. Fictitious paradoxes like that of Einstein, Podolsky, and Rosen[8] which are concerned with such correlated, noninteracting systems are easily investigated and clarified in the present scheme.

Many further combinations of several observers and systems can be studied within the present framework. The results of the present "relative state" formalism agree with those of the conventional "external observation" formalism in all those cases where that familiar machinery is applicable.

In conclusion, the continuous evolution of the state function of a composite system with time gives a complete mathematical model for processes that involve an idealized observer. When interaction occurs, the result of the evolution in time is a superposition of states, each element of which assigns a different state to the memory of the observer. Judged by the state of the memory in almost all of the observer states, the probabilistic conclusion of the usual "external observation" formulation of quantum theory are valid. In other words, pure Process 2 wave mechanics, without any initial probability assertions, leads to all the probability concepts of the familiar formalism.

6. DISCUSSION

The theory based on pure wave mechanics is a conceptually simple, causal theory, which gives predictions in accord with experience. It constitutes a framework in which one can investigate in detail, mathematically, and in a logically consistent manner a number of sometimes puzzling subjects, such as the measuring process itself and the interrelationship of several observers. Objections have been raised in the past to the conventional or "external observation" formulation of quantum theory on the grounds that its probabilistic features are postulated in advance instead of being derived from the theory itelf. We believe that the present "relative-state" formulation meets this objection, while retaining all of the content of the standard formulation.

While our theory ultimately justifies the use of the probabilistic interpretation as an aid to making practical predictions, it forms a broader frame in which to understand the consistency of that interpretation. In this respect it can be said to form a *metatheory* for the standard theory. It transcends the usual "external observation" formulation, however, in its ability to deal logically with questions of imperfect observation and approximate measurement.

The "relative state" formulation will apply to all forms of quantum mechanics which maintain the superposition principle. It may therefore prove a fruitful framework for the quantization of general relativity. The formalism invites one to construct the formal theory first, and to supply the statistical interpretation later. This method should be particularly useful for interpreting quantized unified field theories where there is no question of ever isolating observers and object systems. They all are represented in a *single* structure, the field. Any interpretative rules can probably only be deduced in and through the theory itself.

Aside from any possible practical advantages of the theory, it remains a matter of intellectual interest that the statistical assertions of the usual interpretation do not have the status of independent hypotheses, but are deducible (in the present sense) from the pure wave mechanics that starts completely free of statistical postulates.

[8] Einstein, Podolsky, and Rosen, Phys. Rev. 47, 777 (1935). For a thorough discussion of the physics of observation, see the chapter by N. Bohr in *Albert Einstein, Philosopher-Scientist* (The Library of Living Philosophers, Inc., Evanston, 1949).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Assessment of Everett's "Relative State" Formulation of Quantum Theory

JOHN A. WHEELER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

THE preceding paper puts the principles of quantum mechanics in a new form.[1] Observations are treated as a special case of normal interactions that occur within a system, not as a new and different kind of process that takes place from without. The conventional mathematical formulation with its well-known postulates about probabilities of observations is derived as a *consequence* of the new or "*meta*" quantum mechanics. Both formulations apply as well to complex systems as to simple ones, and as well to particles as to fields. Both supply mathematical models for the physical world. In the new or "relative state" formalism this model associates with an isolated system a state function that obeys a linear wave equation. The theory deals with the totality of all the possible ways in which this state function can be decomposed into the sum of products of state functions for subsystems of the overall system—and nothing more. For example, in a system endowed with four degrees of freedom x_1,x_2,x_3,x_4, and a time coordinate, t, the general state can be written $\psi(x_1,x_2,x_3,x_4,t)$. However, there is *no* way in which ψ defines any unique state for any subsystem (subset of x_1,x_2,x_3,x_4). The subsystem consisting of x_1 and x_3, say, *cannot* be assigned a state $u(x_1,x_3,t)$ independent of the state assigned to the subsystem x_2 and x_4. In other words, there is ordinarily *no* choice of f or u which will allow ψ to be written in the form $\psi=u(x_1,x_3,t)f(x_2,x_4,t)$. The most that can be done is to associate a *relative* state to the subsystem, $u_{\mathrm{rel}}(x_1,x_3,t)$, relative to some *specified* state $f(x_2,x_4,t)$ for the remainder of the system. The method of assigning relative states $u_{\mathrm{rel}}(x_1,x_3,t)$ in one subsystem to specific states $f(x_2,x_4,t)$ for the remainder, permits one to decompose ψ into a superposition of products, each consisting of one member of an orthonormal set for one subsystem and its corresponding relative state in the other subsystem:

$$\psi=\sum_i a_i f_i(x_2,x_4,t)u_{\mathrm{rel}\,f_i}(x_1,x_3,t), \qquad (1)$$

where $\{f\}$ is an orthonormal set. According as the functions f_n constitute one or another family of orthonormal functions, the relative state functions $u_{\mathrm{rel}\,f_n}$ have one or another dependence upon the variables of the remaining subsystem.

Another way of phrasing this unique association of relative state in one subsystem to states in the remainder is to say that the states are correlated. The totality of these correlations which can arise from all possible decompositions into states and relative states is all that can be read out of the mathematical model.

The model has a place for observations only insofar as they take place within the isolated system. The theory of observation becomes a special case of the theory of correlations between subsystems.

How does this mathematical model for nature relate to the present conceptual scheme of physics? Our conclusions can be stated very briefly: (1) The conceptual scheme of "relative state" quantum mechanics is completely different from the conceptual scheme of the conventional "external observation" form of quantum mechanics and (2) The conclusions from the new treatment correspond completely in familiar cases to the conclusions from the usual analysis. The rest of this note seeks to stress this *correspondence in conclusions* but also this *complete difference in concept*.

The "external observation" formulation of quantum mechanics has the great merit that it is dualistic. It associates a state function with the system under study —as for example a particle—but not with the *ultimate* observing equipment. The system under study can be enlarged to include the original object as a subsystem and also a piece of observing equipment—such as a Geiger counter—as another subsystem. At the same time the number of variables in the state function has to be enlarged accordingly. However, the *ultimate* observing equipment still lies outside the system that is treated by a wave equation. As Bohr[2] so clearly emphasizes, we always interpret the wave amplitude by way of observations of a classical character made from outside the quantum system. The conventional formalism admits no other way of interpreting the wave amplitude; it is logically self-consistent; and it rightly rules out any classical description of the internal dynamics of the system. With the help of the principle of complementarity the "external observation" formulation nevertheless keeps all it consistently can of classical concepts. Without this possibility of classical measuring equipment the mathematical machinery of quantum mechanics would seem at first sight to admit no correlation with the physical world.

Instead of founding quantum mechanics upon classical physics, the "relative state" formulation uses a completely different kind of model for physics. This new model has a character all of its own; is conceptually

[1] Hugh Everett, III, Revs. Modern Phys. 29, 454 (1957).

[2] Chapter by Niels Bohr in *Albert Einstein, Philosopher-Scientist*, edited by P. A. Schilpp (The Library of Living Philosophers, Inc., Evanston, Illinois, 1949).

self-contained; defines its own possibilities for interpretation; and does not require for its formulation any reference to classical concepts. It is difficult to make clear how decisively the "relative state" formulation drops classical concepts. One's initial unhappiness at this step can be matched but few times in history[3]: when Newton described gravity by anything so preposterous as action at a distance; when Maxwell described anything as natural as action at a distance in terms as unnatural as field theory; when Einstein denied a privileged character to any coordinate system, and the whole foundations of physical measurement at first sight seemed to collapse. How can one consider seriously a model for nature that follows neither the Newtonian scheme, in which coordinates are functions of time, nor the "external observation" description, where probabilities are ascribed to the possible outcomes of a measurement? Merely to analyze the alternative decompositions of a state function, as in (1), without saying what the decomposition means or how to interpret it, is apparently to define a theoretical structure almost as poorly as possible! Nothing quite comparable can be cited from the rest of physics except the principle in general relativity that all regular coordinate systems are equally justified. As in general relativity, so in the relative-state formulation of quantum mechanics the analysis of observation is the key to the physical interpretation.

Observations are not made from outside the system by some super-observer. There is no observer on hand to use the conventional "external observation" theory. Instead, the whole of the observer apparatus is treated in the mathematical model as part of an isolated system. All that the model will say or ever can say about observers is contained in the interrelations of eigenfunctions for the object part of this isolated system and relative state functions of the remaining part of the system. Every attempt to ascribe probabilities to observables is as out of place in the relative state formalism as it would be in any kind of quantum physics to ascribe coordinate and momentum to a particle at the same time. The word "probability" implies the notion of observation from outside with equipment that will be described typically in classical terms. Neither these classical terms, nor observation from outside, nor a priori probability considerations come into the *foundations* of the relative state form of quantum theory.

So much for the conceptual differences between the new and old formulations. Now for their correspondence. The preceding paper shows that this correspondence is detailed and close. The tracing out of the correspondence demands that the system include something that can be called an observing subsystem. This subsystem can be as simple as a particle which is to collide with a particle that is under study. In this case the correspondence occurs at a primitive level between the relative state formalism where the system consists of two particles, and the external observation theory where the system consists of only one particle. The correlations between the eigenfunctions of the object particle and the relative state functions of the observer particle in the one scheme are closely related in the other scheme to the familiar statements about the relative probabilities for various possible outcomes of a measurement on the object particle.

A more detailed correspondence can be traced between the two forms of quantum theory when the observing system is sufficiently complex to have what can be described as memory states. In this case one can see the complementary aspects of the usual external observation theory coming into evidence in another way in the relative state theory. They are expressed in terms of limitations on the degree of correlation between the memory states for successive observations on a system of the same quantity, when there has been an intervening observation of a noncommuting quantity. In this sense one has in the relative state formalism for the first time the possibility of a closed mathematical model for complementarity.

In physics it is not enough for a single observer or apparatus to make measurements. Different pieces of equipment that make the same type of measurement on the same object system must show a pattern of consistency if the concept of measurement is to make sense. Does not such consistency demand the external observation formulation of quantum theory? There the results of the measurements can be spelled out in classical language. Is not such "language" a prerequisite for comparing the measurements made by different observing systems?

The analysis of multiple observers in the preceding paper by the theory of relative states indicates that the necessary consistency between measurements is already obtained without going to the external observer formulation. To describe this situation one can use if he will the words "communication in clear terms always demands classical concepts." However, the kind of physics that goes on does not adjust itself to the available terminology; the terminology has to adjust itself in accordance with the kind of physics that goes on. In brief, the problem of multiple observers solves itself within the theory of relative states, not by adding the conventional theory of measurement to that theory.

It would be too much to hope that this brief survey should put the relative state formulation of quantum theory into completely clear focus. One can at any rate end by saying what it does not do. It does not seek to supplant the conventional external observer formalism, but to give a new and independent foundation for that formalism. It does not introduce the idea of a super-observer; it rejects that concept from the start. It does not supply a prescription to say what is the correct

[3] See, for example, Philipp Frank's *Modern Science and Its Philosophy* (George Braziller, New York, 1955), Chap. 12.

functional form of the Hamiltonian of any given system. Neither does it supply any prediction as to the functional dependence of the over-all state function of the isolated system upon the variables of the system. But neither does the classical universe of Laplace supply any prescription for the original positions and velocities of all the particles whose future behavior Laplace stood ready to predict. In other words, the relative state theory does not pretend to answer all the questions of physics. The concept of relative state does demand a totally new view of the foundational character of physics. No escape seems possible from this relative state formulation if one wants to have a complete mathematical model for the quantum mechanics that is internal to an isolated system. Apart from Everett's concept of relative states, no self-consistent system of ideas is at hand to explain what one shall mean by quantizing[4] a closed system like the universe of general relativity.

[4] C. W. Misner, Revs. Modern Phys. **29**, 497 (1957).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Interaction of Neutrinos and Gravitational Fields

DIETER R. BRILL AND JOHN A. WHEELER

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

1. INTRODUCTION; GRAVITATION THE ONLY FORCE IN WHICH NEUTRINOS ARE SUBJECT TO SIMPLE ANALYSIS

KNOWLEDGE of neutrinos to date is confined mainly to emission and absorption processes; that is, to the domain of elementary particle transformations. For comparison, imagine that one knew about electrons only the rate at which they are produced in beta decay, or absorbed in K-electron absorption processes, but knew nothing about the motion of electrons in electric and magnetic fields, nothing about the binding of electrons in atoms or the existence of spin-orbit coupling and very little about the stress energy tensor of the electron. What can one do to learn some fraction as much about neutrinos as one knows today about electrons?

The neutrino does not respond directly to electric or magnetic fields. Therefore, if one wishes to influence its orbit by forces subject to simple analysis one has to make use of gravitational fields. In other words, one has to consider the physics of a neutrino in a curved metric.

For this task the only available tools of analysis are theoretical. We accept the recently clarified[1] and dramatically tested[2,3] neutrino theory. We see no motive to change the theory. Instead we recall in Sec. 2 the clearly defined extension of the Dirac equation to the curved space that represents the most general gravitational field. In Sec. 3, we specialize to the neutrino with its zero mass and to the class of solutions with right-handed polarization that are demanded by the recently gained knowledge.[1–3] Section 4 separates out the radial wave equation for the motion of a neutrino in a centrally symmetric gravitational field, and identifies one term in this equation with a spin-orbit coupling. Section 5 compares and contrasts the energy level spectrum in the case of spherical symmetry for (1) an electron in an electrostatic field, (2) an electron in a gravitational field, (3) a photon in a gravitational field, and (4) a neutrino in a gravitational field. Section 6 recalls the statistical mechanics of an ensemble of neutrinos. Section 7 discusses some neutrino pair creation processes that do not depend upon beta interactions for their existence. Section 8 deals with the contribution of neutrinos to the stress energy tensor, Sec. 9 deals with the gravitational interaction of two neutrinos traveling parallel or antiparallel to each other; and Sec. 10 with the contribution to the stress energy tensor due to a neutrino in a bound orbit. Finally, Sec. 11 examines by way of illustration an object where both the creation of gravitational fields by neutrinos, and the response of neutrinos to gravitational fields come into play: a geon or entity constituted entirely of neutrinos and held together by their mutual gravitational attractions.

2. MATHEMATICS OF SPIN IN CURVED SPACE

Spinor fields have been treated in general relativity by many authors[4] and from three principal points of view (Table I). The three formalisms are in principle equivalent and must therefore in any actual problem give identical results for such well-defined quantities as

[1] T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956); **105**, 1119(L) (1957).

[2] Wu, Ambler, Hayward, Hoppes, and Hudson, Phys. Rev. **105**, 1413(L) (1957).

[3] Garwin, Lederman, and Weinrich, Phys. Rev. **105**, 1415(L) (1957).

[4] See W. L. Bade and H. Jehle, Revs. Modern Phys. **25**, 714 (1953) for a general review of the literature. To their list of principal references one should add M. Riesz, Lund Univ. Math. Sem. Band 12 (1954); F. J. Belinfante, Physica **7**, 305 (1940).

TABLE I. Comparison of the three principal formalisms for description of spin in general relativity.

	General formalism		Vierbein formalism	
Components of ψ	4	2	4	2
Fundamental spin matrices $(i=1,\cdots,4)$	γ_i	s_i	$\bar{\gamma}_i$	$\bar{s}_i$
Dependent on position	yes	yes	no	no
Relation between spin matrices and metric	Spin matrices conform to metric $[\gamma_i,\gamma_j]_+=2g_{ij}$	$[\hat{s}_i,s_j]_+=2g_{ij}$	Metric transformed to locally Lorentz metric $d\bar{x}^k=b_\alpha{}^k dx^\alpha$ $g_{ik}=b_i{}^\alpha b_k{}^\beta g_{\alpha\beta}$ Lorentz $\gamma_k=b_k{}^\alpha\bar{\gamma}_\alpha$ Use flat space spin matrices	$s_k=b_k{}^\alpha\bar{s}_\alpha$
Covariant under	General coordinate and similarity transformations		General coordinate transformations, and quite independent Lorentz transformations of the Vierbein differentials, $d\bar{x}^k$	
Most general form for second of these transformations	16 parameters	4 complex parameters	6 real Lorentz parameters, with or without inversions of space or time or both.	
Formation of covariants	As in tensor analysis; see Table II for covariant derivative			
Pauli conjugate, $\psi^\dagger$	$\psi^*\cdot i\cdot$(Hermitizing matrix)		$\psi^* i\bar{\gamma}^4$	$\psi^* i\bar{s}^4$comp. conj.
Dirac equation for neutrino	$\gamma^\alpha\nabla_\alpha\psi=0$	$s^\alpha\nabla_\alpha\psi=0$	$\gamma^\alpha\nabla_\alpha\psi=0$	$s^\alpha\nabla_\alpha\psi=0$
Lagrangian, L	$\psi^\dagger\gamma^\alpha\nabla_\alpha\psi$	$\psi^\dagger s^\alpha\nabla_\alpha\psi$	$\psi^\dagger\gamma^\alpha\nabla_\alpha\psi$	$\psi^\dagger s^\alpha\nabla_\alpha\psi$
Current 4-vectors, s^k	$\psi^\dagger i\gamma^k\psi$	$\psi^\dagger is^k\psi$	$\psi^\dagger i\gamma^k\psi$	$\psi^\dagger is^k\psi$
Vierbein formalism for Dirac equation in polar coordinates for two simple choices of Vierbeine				
Vierbeine parallel to unit vectors in	r,θ,φ,T directions		x,y,z,T directions	
Corresponding γ matrices	$\gamma_r=(\exp\frac{1}{2}\lambda)\bar{\gamma}_1$ $\gamma_\theta=r\bar{\gamma}_2$ $\gamma_\varphi=r\sin\theta\bar{\gamma}_3$ $\gamma_r=(\exp\frac{1}{2}\nu)\bar{\gamma}_4$		$\gamma_r=(\exp\frac{1}{2}\lambda)(\sin\theta\sin\varphi\bar{\gamma}_1+\sin\theta\cos\varphi\bar{\gamma}_2+\cos\theta\bar{\gamma}_3)$; similarly for γ_θ and γ_φ; $\gamma_T=(\exp\frac{1}{2}\nu)\bar{\gamma}_4$	
Separation of Dirac wave function for simple choice of representation of the γ's	$\begin{pmatrix}F(r)f(\theta)\\F(r)g(\theta)\\G(r)f(\theta)\\G(r)g(\theta)\end{pmatrix}\exp(im_j\varphi-iEt/h)$		$\begin{pmatrix}F(r)Y_1(\theta)\exp[i(m_j+\frac{1}{2})\varphi-iEt/\hbar]\\F(r)Y_2(\theta)\exp[i(m_j-\frac{1}{2})\varphi-iEt/\hbar]\\G(r)Y_1(\theta)\exp[i(m_j+\frac{1}{2})\varphi-iEt/\hbar]\\G(r)Y_2(\theta)\exp[i(m_j-\frac{1}{2})\varphi-iEt/\hbar]\end{pmatrix}$	
Spinor wave function is a single valued function of position	no		yes	

energy eigenvalues and density of stress and energy. We find it convenient to use here the formalism of V. Bargmann because of its generality. The equations of this formalism are covariant with respect to general coordinate transformations, and invariant under general, i.e., position-dependent, similarity transformations of the spinors. The fundamental connection between space and spin is made through a field of γ matrices which satisfy the anticommutation relationship,

$$\gamma_i\gamma_k+\gamma_k\gamma_i=2g_{ik}1, \tag{1}$$

at each point in space, where g_{ij} is the metric tensor at that point, and 1 is the unit matrix. For the *signature* of the metric tensor we adopt the familiar Pauli convention $(1, 1, 1, -1)$. Let the dependence of the g_{ik} upon position be known. Then one can set up a generally covariant spinor formalism with the help of any field of 4×4 matrices γ_i that have the following properties: (1) their components are continuous functions of position in space time, (2) they satisfy (1) and (3) they transform like a vector under coordinate transformations, (4) under the spinor transformation

$$(\text{spinor})_{\text{new}}=S^{-1}(\text{spinor})_{\text{old}},$$

they undergo the similarity transformation

$$(\gamma_i)_{\text{new}}=S^{-1}(\gamma_i)_{\text{old}}S.$$

It also substantially simplifies the treatment of charged spinor fields to limit attention to real representations, γ_k, of the spin matrices, and to limit attention to spinor transformations, S, whose matrix elements are also real.

The principal feature of the mathematical formalism is a definition of covariant differentiation which is the natural generalization of the covariant differentiation of tensor analysis (Table II). In addition to the usual Christoffel symbols, $\Gamma_{ik}{}^m$, formed from the metric, g_{ik}, it is necessary to introduce four 4×4 matrices Γ_k. These quantities are uniquely determined up to an

TABLE II. Covariant differentiation of spin dependent quantities compared and contrasted with covariant differentiation of tensors.

	Tensors	Quantities with spinor transformation properties
Symbol for covariant differentiation with respect to x^k	Final subscript $;k$	Preceding operator ∇_k
Special form when space is flat, coordinates are Euclidean, and γ's are independent of position	$\partial/\partial x^k$	$\partial/\partial x^k$
Additional quantities needed to define covariant derivative when one or more of these conditions are not fulfilled	40 functions of position $\Gamma_{ik}{}^m$	The $\Gamma_{ik}{}^m$ and four matrices, Γ_k
Formula to determine these additional quantities from the metric	$\frac{1}{2}g^{m\alpha}\left(\frac{\partial g_{k\alpha}}{\partial x^i}+\frac{\partial g_{i\alpha}}{\partial x^k}-\frac{\partial g_{ik}}{\partial x^\alpha}\right)$	Eq. (2) or Eq. (8)
How these quantities enter into the definition of the covariant derivative	Depends upon the transformation properties of the quantity being differentiated	
Example I	Scalar, f	Spinor,
Effect of a coordinate transformation	$f_{\text{new}}=f_{\text{old}}$	$\psi_{\text{new}}=\psi_{\text{old}}$
Effect of a spinor transformation	$f_{\text{new}}=f_{\text{old}}$	$\psi_{\text{new}}=S^{-1}\psi_{\text{old}}$
Covariant derivative	$f_{;k}=\partial f/\partial x^k$	$\nabla_k\psi=\partial\psi/\partial x^k-\Gamma_k\psi$
Example II	Vector, A^s	Conjugate spinor $\psi^\dagger$
Effect of a coordinate transformation	$A^s{}_{\text{new}}=\frac{\partial\bar{x}^s}{\partial x^\alpha}A^\alpha{}_{\text{old}}$	$\psi_{\text{new}}{}^\dagger=\psi_{\text{old}}{}^\dagger$
Effect of a spinor transformation	$A^s{}_{\text{new}}=A^s{}_{\text{old}}$	$\psi_{\text{new}}{}^\dagger=\psi_{\text{old}}{}^\dagger S$
Covariant derivative	$A^s{}_{;k}=\frac{\partial A^s}{\partial x^k}+\Gamma_{k\alpha}{}^s A^\alpha$	$\nabla_k\psi^\dagger=\partial\psi^\dagger/\partial x^k+\psi^\dagger\Gamma_k$
Example III	Tensor, $M_i{}^k$	Spinor tensor, T_{ik} (such as $\gamma_i\gamma_k$)
Effect of a coordinate transformation	$M_i{}^k{}_{\text{new}}=\frac{\partial\bar{x}^k}{\partial x^\alpha}\frac{\partial x^\beta}{\partial\bar{x}^i}M_\alpha{}^\beta{}_{\text{old}}$	$T_{ik\text{ new}}=\frac{\partial x^\alpha}{\partial\bar{x}^i}\frac{\partial x^\beta}{\partial\bar{x}^k}T_{\alpha\beta\text{ old}}$
Effect of a spinor transformation	$M_i{}^k{}_{\text{new}}=M_i{}^k{}_{\text{old}}$	$T_{ik\text{ new}}=S^{-1}T_{ik\text{ old}}S$
Covariant derivative	$M_i{}^k{}_{;l}=\frac{\partial M_i{}^k}{\partial x^l}+\Gamma_{l\alpha}{}^k M_i{}^\alpha-\Gamma_{il}{}^\alpha M_\alpha{}^k$	$\nabla_l T_{ik}=T_{ik;l}+T_{ik}\Gamma_k-\Gamma_k T_{ik}$

additive multiple of the unit matrix by

$$\partial\gamma_i/\partial x^k-\Gamma_{ik}{}^\mu\gamma_\mu-\Gamma_k\gamma_i+\gamma_i\Gamma_k=0. \tag{2}$$

The $\Gamma_{ik}{}^m$ and Γ_k together permit one to define the covariant derivative of any object of which the transformation properties for general coordinate and similarity transformations are known. This covariant derivative is denoted by ∇_k; its explicit form depends on the quantity it acts on (Table II). This covariant differentiation has the standard properties

$$\begin{aligned}\nabla_k(AB)&=(\nabla_k A)B+B(\nabla_k A),\\ \nabla_k(A^*)&=(\nabla_k A)^*,\\ \nabla_k\gamma_i&=0,\end{aligned} \tag{3}$$

where the symbol * means Hermitian adjoint (the transpose of the complex conjugate). The quantity, $\nabla_k A$, transforms like A under similarity transformations and like a tensor of one higher rank under coordinate transformations.

The definition of the γ_i by (1) can always be fulfilled by a linear combination of the matrices of special relativity. These are constant matrices—and if one wishes, purely real matrices—that satisfy the conditions

$$\tilde{\gamma}_i\tilde{\gamma}_k+\tilde{\gamma}_k\tilde{\gamma}_i=2g_{ik\text{ Lorentz}},$$

$$g_{ik\text{ Lorentz}}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&-1\end{pmatrix}, \tag{4}$$

$$\tilde{\gamma}_i{}^*=\tilde{\gamma}_i\ \ (i=1,2,3),\quad \tilde{\gamma}_4{}^*=-\tilde{\gamma}_4. \tag{5}$$

Thus, at each point, x, in 4 space it is possible to transform from the general coordinates x^i to a system $\tilde{x}^i$ ("Vierbein") whose metric is Minkowskian at that point:

$$dx^k=a^k{}_\alpha d\tilde{x}^\alpha,\quad d\tilde{x}^k=b_\beta{}^k dx^\beta. \tag{6}$$

Then (1) is satisfied by

$$\gamma_i=b_i{}^\alpha\tilde{\gamma}_\alpha. \tag{7}$$

The formalism which confines itself to solutions of (1) of the special type (7) is called Vierbein formalism. In terms of the Vierbein components $b_i{}^j$, $a^k{}_l$ the explicit solution of (2) has the simple form[5]

$$\Gamma_k = g_{\mu\alpha}[(\partial b_\nu{}^\beta/\partial x^k)a^\alpha{}_\beta - \Gamma_{\nu l}{}^\alpha]s^{\mu\nu} + a_k 1, \tag{8}$$

where $s^{ij}=\frac{1}{2}(\gamma^i\gamma^j-\gamma^j\gamma^i)$ and a_k is arbitrary. In the Vierbein frame a similarity transformation of the spinors is equivalent to a Lorentz transformation of the Vierbein. Thus the invariance of the formalism under similarity transformations can be understood geometrically as invariance under arbitrary Lorentz transformations of the Vierbein. These Lorentz transformations have nothing to do with any coordinate transformation, and vary arbitrarily from point to point in space.

The Dirac equation in general relativity can be written in the form

$$\gamma^\alpha \nabla_\alpha \psi + \mu\psi = 0 \quad (\mu = mc/\hbar). \tag{9}$$

Here we assume the arbitrary traces of the Γ_k have been adjusted so as to account for the effects of the electromagnetic potentials.

To form real expressions, as for example for the current density, we can follow Bargmann with minor differences in notation and define the "Pauli conjugate" of ψ,

$$\psi^\dagger = \psi^* \alpha$$

by means of a "Hermitizing matrix," α, chosen so that both α itself and the four matrices, $\alpha i\gamma^k$, are Hermitian. In the Vierbein formalism we may choose the $\tilde{\gamma}^k$ to be real and choose[6] $\alpha = i\tilde{\gamma}^4$. One then can form a current density,

$$s^k = \psi^\dagger i\gamma^k\psi, \tag{10}$$

whose covariant divergence vanishes by virtue of the Dirac equation (9). The Dirac equation can be derived from the variational principle

$$\delta\int L(-g)^{\frac{1}{2}}d^4x = 0, \tag{11}$$

with

$$L = \psi^\dagger\gamma^\alpha(\nabla_\alpha\psi) + \mu\psi^\dagger\psi, \tag{12}$$

by independent variation of $\psi^\dagger$ and ψ.

3. ALPHA AND BETA ROTATIONS AND TWO-COMPONENT NEUTRINO WAVE EQUATION

We set the mass term equal to zero in the Dirac equation (9) and the Lagrangian (12) to obtain the wave equation for a Dirac-type neutrino,

$$\gamma^\alpha\nabla_\alpha\psi = 0. \tag{13}$$

In this special case of zero rest-mass, the wave equation and the Lagrangian are invariant under a wider class of transformations than are usually considered in the Dirac theory. The situation is analogous to the case of the charge-free electromagnetic field, F_{ik}. Consider any solution of Maxwell's equations,

$$\begin{aligned} \epsilon^{i\alpha\beta\gamma}\partial F_{\alpha\beta}/\partial x^\gamma &= 0, \\ (-g)^{-\frac{1}{2}}(\partial/\partial x^\alpha)(-g)^{\frac{1}{2}}F^{\alpha k} &= 0, \end{aligned} \tag{14}$$

where $\epsilon^{1234}=1$, and ϵ^{ijkl} changes sign on permutation of any two indices. From any solution of these equations in a metric of arbitrary curvature one can generate a new solution by a special kind of transformation that we may call an "α-rotation";

$$(F_{ik})_{\text{new}} = F_{ik}\cos\alpha + \tfrac{1}{2}(-g)^{-\frac{1}{2}}g_{i\sigma}g_{k\tau}\epsilon^{\sigma\tau\alpha\beta}F_{\alpha\beta}\sin\alpha, \tag{15}$$

where α is an angle that is independent of position and time. In flat space this transformation takes the form

$$\begin{aligned} \mathbf{E}_{\text{new}} &= \mathbf{E}\cos\alpha + \mathbf{H}\sin\alpha, \\ \mathbf{H}_{\text{new}} &= \mathbf{H}\cos\alpha - \mathbf{E}\sin\alpha. \end{aligned} \tag{16}$$

It is plainly not a rotation in any ordinary sense except in the special case where $\mathbf{E}$ and $\mathbf{H}$ represent the principal polarization directions of a monochromatic plane wave; then the α transformation rotates the axes of polarization by the angle α.

Similarly, let ψ be a solution of the Dirac equation for zero mass in a space-time continuum of arbitrary curvature. Then one can generate a new solution by the "β rotation,"[7]

$$\psi_{\text{new}} = \exp(\tfrac{1}{2}\beta\gamma_5)\psi = [1\cos(\tfrac{1}{2}\beta) + \gamma_5\sin(\tfrac{1}{2}\beta)], \tag{17}$$

where β is a constant and

$$\begin{aligned} \gamma_5 &= (-g)^{-\frac{1}{2}}(1/4!)\epsilon^{\alpha\beta\gamma\delta}\gamma_\alpha\gamma_\beta\gamma_\gamma\gamma_\delta, \\ (\gamma_5)^2 &= -1, \\ \gamma_5\gamma_i &= -\gamma_i\gamma_5 \quad (i=1,2,3,4). \end{aligned} \tag{18}$$

In the special case where the neutrino wave is linearly polarized in a given region of space, the β rotation turns this direction of polarization through the angle β. That $\gamma_5\psi$ is a solution of (13) follows from (3) and the anticommutation relations (18); hence the linear combination, (17), of ψ and $\gamma_5\psi$ is also a solution.

No change whatever in the metric can lift the degeneracy between the spin polarization states ψ and $\exp(\frac{1}{2}\beta\gamma_5)\psi$. An analogous situation occurs in the physics of a two electron system. No allowable system of forces can ever produce a difference in energy between the states $u(x_1,x_2)$ and $\exp(i\gamma P_{12})u(x_1,x_2)$, where P_{12} is the permutation operator. Nature apparently does not ever permit an irrevocable degeneracy of this kind. Only the combination $u(x_1,x_2)-u(x_2,x_1)$ is allowed. Assume similarly that nature rules out a duplicity of spin states for the neutrino that could never be split by any gravita-

[5] For the solution for general γ fields, see O. Klein, Arch. Math., Astr. och Phys. 34, 1 (1947).

[6] For the general Hermitizing matrix, see W. Kofink, Math. Z. 51, 702 (1949).

[7] The possibility of this transformation in *flat* space is well known.

tional field, however strong. More specifically, assume that the only allowed state, ψ, for the neutrino is a state that is transformed into a multiple of itself by every β rotation:

$$\exp(\tfrac{1}{2}\beta\gamma_5)\psi_{\text{allowed}}=e^{i\,(\text{constant})}\psi_{\text{allowed}}. \qquad (19)$$

Then one concludes that the allowed state functions necessarily have *circular polarization*, in the sense that the expression

$$\psi_{\text{allowed}}=\psi_c=(1+i\gamma_5)\psi, \qquad (20)$$

constitutes a mixture with 90° phase difference of the states with rotations of 0° and 180°. To change the sign of i in (20) is only to interchange the definitions of positive and negative energy states. Lee and Yang[1] have recently given different arguments for considering all neutrinos to have right-handed circular polarization. Their considerations have received dramatic verification.[2,3] The conclusion appears inescapably that neutrinos possess only one state of polarization, which is circular.

The allowable spinor state functions satisfy the condition

$$(1-i\gamma_5)\psi_c=0, \qquad (21)$$

and in a suitable representation have only two nonzero components. They can be described by Pauli's spinors of two components, as Lee and Yang show. They introduce the two-component wave equation

$$H\psi=c\boldsymbol{\sigma}\cdot\mathbf{p}\psi=i\hbar\partial\psi/\partial t, \qquad (22)$$

where $\boldsymbol{\sigma}$ are the three 2×2 Pauli spin matrices. To write this equation in generally covariant form, it is convenient to introduce four 2×2 matrices s_i which satisfy the conditions

$$[\bar{s}_i,s_j]_+\equiv\bar{s}_is_j+\bar{s}_js_i=2g_{ij}, \qquad (23)$$

where the bar denotes complex conjugation. Then the covariant form of the Pauli-Lee-Yang equation is

$$s^\alpha\nabla_\alpha\psi=0. \qquad (24)$$

The correct interpretation of the s^i and the covariant derivative ∇_i is well known from spinor analysis.[8]

A beta rotation is not the only means to generate a new solution of the neutrino wave equation (13) from a general solution, ψ. Let a representation be employed in which the basic matrices are all real, $\gamma^i{}_{\text{complex conjugate}}=\gamma^i$, and let the complex conjugate of (13) be taken; then it follows at once that $\psi_{\text{complex conjugate}}$ satisfies the wave equation as well as does ψ itself. When ψ represents a positive energy state, $\psi_{\text{complex conjugate}}$ of course represents a negative energy state; but there is ordinarily no well-defined distinction between the two kinds of states in a metric which varies both with space and with time.

How does one know that the basic matrices can still be taken real when the three-dimensional space under consideration has an arbitrary curvature and topolog[y]? We assume for simplicity that time has the topology [of] a straight line. We are indebted to Professor V. Bar[g]mann for informing us of a theorem cited by Hopf [at] the International Congress of Mathematicians, Cam[bridge], Massachusetts (Vol. I, p. 193 of the 19[50] Proceedings) to the effect that one can always in [a] *three*-dimensional space define three mutually orthogon[al] nonsingular vector fields—a construction that is we[ll] known to be impossible on the *two*-dimensional close[d] surface of a sphere. Then one has only to take over t[he] standard four *real* $\tilde{\gamma}$ matrices of flat space and emplo[y] the Vierbein formalism to have four real matrices,

$$\gamma_i=b_i{}^\alpha\tilde{\gamma}_\alpha, \qquad (2\text{-}$$

in the curved space under consideration.

4. MOTION OF A NEUTRINO IN A SPHERICALLY SYMMETRICAL GRAVITATION FIELD

We explore the reaction of the neutrino to a gravita[]tional field in the simplest known case: a sphericall[y] symmetric metric of the Schwarzschild type,

$$g_{ij}=\begin{bmatrix} e^\lambda & 0 & 0 & 0\\ 0 & r^2 & 0 & 0\\ 0 & 0 & r^2\sin^2\theta & 0\\ 0 & 0 & 0 & -e^\nu\end{bmatrix}. \qquad (2[6])$$

Here the coordinates are $x^1=r$, $x^2=\theta$, $x^3=\varphi$, $x^4=ct=T$ The dilation functions $\lambda(r)$ and $\nu(r)$ are assumed t[o] depend upon distance in an arbitrary way; they are no[t] limited to the special Schwarzschild solution for [a] localized concentration of mass,

$$e^{-\lambda}=e^\nu=1-2GM/c^2r. \qquad (2[7])$$

In order to write the Dirac equation in this metric w[e] choose a field of γ matrices of the type (7). Two choice[s] for these matrices are simple: (1) Vierbein axes parall[el] to the r, θ, and φ axes at each point, so that the desire[d] Dirac matrices, γ^i are expressed in terms of the standar[d] Dirac matrices, $\tilde{\gamma}^i$, for a Cartesian coordinate systen[m] by the formulas

$$\gamma_1=e^{\frac{1}{2}\lambda}\tilde{\gamma}_1,\quad \gamma_2=r\tilde{\gamma}_2,\quad \gamma_3=r\sin\theta\tilde{\gamma}_3,\quad \gamma_4=e^{\frac{1}{2}\nu}\tilde{\gamma}_4, \qquad (2[8])$$

and (2), Vierbein axes parallel to some rectangular co[]ordinate system:

$$\begin{aligned}\gamma_1&=e^{\frac{1}{2}\lambda}(\sin\theta\cos\varphi\tilde{\gamma}_1+\sin\theta\sin\varphi\tilde{\gamma}_2+\cos\theta\tilde{\gamma}_3),\\ \gamma_2&=r(\cos\theta\cos\varphi\tilde{\gamma}_1+\cos\theta\sin\varphi\tilde{\gamma}_2-\sin\theta\tilde{\gamma}_3),\\ \gamma_3&=r\sin\theta(-\sin\varphi\tilde{\gamma}_1+\cos\varphi\tilde{\gamma}_2),\\ \gamma_4&=e^{\frac{1}{2}\nu}\tilde{\gamma}_4.\end{aligned} \qquad (2[9])$$

The two choices lead to the same radial equation, bu[t] to a different dependence of the components of th[e] spinor wave function upon the angles (Table I). I[n] case (2) the angular dependence agrees with that pr[e]dicted in the familiar case of special relativity b[]

[8] See, for example, W. L. Bade and H. Jehle, Revs. Modern Phys. 25, 714 (1953).

standard formulas.[9] Such ambiguities in the wave function are to be expected whenever one uses a formalism which is invariant under similarity transformations that are quite independent of coordinate transformations. The physically meaningful quantities, such as the current density, will, of course, agree for both choices, for the similarity transformation leaves these quantities invariant. We show the explicit calculations for our choice 1. The Γ_k are found from (2) or (8),

$$\begin{aligned}&\Gamma_1=0;\quad \Gamma_2=\tfrac{1}{2}e^{-\frac{1}{2}\lambda}\tilde{\gamma}_1\tilde{\gamma}_2;\\&\Gamma_3=\tfrac{1}{2}(\sin\theta e^{-\frac{1}{2}\lambda}\tilde{\gamma}_1\tilde{\gamma}_3+\cos\theta\tilde{\gamma}_2\tilde{\gamma}_3);\\&\Gamma_4=(\nu'/4)e^{\frac{1}{2}(\nu-\lambda)}\tilde{\gamma}_1\tilde{\gamma}_4,\end{aligned}\tag{30}$$

where $\nu'=d\nu/dr$.

The Dirac equation for a particle of mass $\mu\hbar/c$ takes the form

$$\begin{aligned}0&=\gamma^\alpha\left(\frac{\partial}{\partial x^\alpha}-\Gamma_\alpha\right)\psi+\mu\psi\\&=[\tilde{\gamma}_1e^{-\frac{1}{2}\lambda}(\partial/\partial r+r^{-1}+\nu'/4)+(\tilde{\gamma}_2/r)(\partial/\partial\theta+\tfrac{1}{2}\cot\theta)\\&\quad+(\tilde{\gamma}_3/r\sin\theta)(\partial/\partial\varphi)-e^{-\frac{1}{2}\nu}\tilde{\gamma}_4(\partial/\partial T)]\psi+\mu\psi.\end{aligned}\tag{31}$$

We can rewrite this equation, following Schrödinger, as

$$i\partial\omega/\partial T=h\omega,\tag{32}$$

with

$$\omega=\exp(\nu/4)r(\sin\theta)^{\frac{1}{2}}\psi,\tag{33}$$

and

$$e^{-\frac{1}{2}\nu}h=\tilde{\gamma}_4\tilde{\gamma}_1e^{-\frac{1}{2}\lambda}\partial/i\partial r+(\tilde{\gamma}_1/r)K-i\tilde{\gamma}_4\mu,\tag{34}$$

$$K=\tilde{\gamma}_1\tilde{\gamma}_4\tilde{\gamma}_2\partial/i\partial\theta+\tilde{\gamma}_1\tilde{\gamma}_4\tilde{\gamma}_3\partial/i\sin\theta\partial\varphi.\tag{35}$$

The operator K is Hermitian and commutes with h. We can therefore choose simultaneous eigenfunctions of h and K, and separate the μth component of the wave function, ω_μ, into radial, angular and time factors,

$$\omega_\mu=R_\mu(r)\Theta_\mu(\theta,\varphi)\exp(-ihT).\tag{36}$$

Here the quantity h represents (energy/$\hbar c$). The angular factor, Θ, is determined by the requirement

$$K\omega=k\omega,\tag{37}$$

where k is constant. This equation for eigenstates of the angular motion has been investigated by Schrödinger.[10] He finds, in agreement with the conventional treatment of the Dirac equation in a central field, a spectrum of positive and negative integral eigenvalues k.

Only the two matrices $\tilde{\gamma}_4$ and $\tilde{\gamma}_1$ remain explicitly in the radial equation (34) after the operator K is replaced by the number k. They can therefore be represented by 2×2 matrices, and the radial factor by a two-component spinor

$$\tilde{\gamma}_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad\tilde{\gamma}_4=\begin{pmatrix}i&0\\0&-i\end{pmatrix},\quad R=\begin{pmatrix}F\\G\end{pmatrix}.\tag{38}$$

In this representation the radial equation for an electron in a Schwarzschild metric, with an electric potential energy $V=vhc$, becomes a modification,

$$\begin{aligned}&[e^{-\frac{1}{2}\nu}(h-v)+\mu]F-e^{-\frac{1}{2}\lambda}dG/dr-(k/r)G=0,\\&[e^{-\frac{1}{2}\nu}(h-v)-\mu]G+e^{-\frac{1}{2}\lambda}dF/dr-(k/r)F=0,\end{aligned}\tag{39}$$

of the familiar radial equations[11] for an electron in a centrally symmetric potential. For the neutrino, of course, we annul both v and μ.

5. COMPARISON OF ENERGY LEVELS OF ELECTRON IN ELECTROSTATIC AND GRAVITATIONAL FIELDS AND NEUTRINO AND PHOTON IN GRAVITATIONAL FIELD

In order to gain some qualitative understanding of the behavior of electrons and neutrinos in gravitational fields we first consider the case of an electron in a gravitational field. We take over the Schwarzschild solution (27) for the metric outside a mass M. We substitute this metric into the radial wave equation (39) and neglect terms of order $1/c^2$ and higher. We find $(E-V)=\hbar c(h-v)$ is replaced by $(E-V-\phi)$, where $\phi=-(GM/r)(E/c^2)$ is the "gravitational potential energy" of a particle of energy E. Therefore in this approximation the energy levels of an electron of positive energy remain unchanged when the electric potential is replaced by an equally strong gravitational potential. The Bohr formulas for energy levels and radii of circular orbits,

$$\begin{aligned}&E=mc^2-\tfrac{1}{2}m(Ze^2/n\hbar)^2,\\&r=n^2\hbar^2/mZe^2,\end{aligned}\tag{40}$$

are replaced by the corresponding formulas

$$\begin{aligned}&E=mc^2-\tfrac{1}{2}m(GMm/n\hbar)^2,\\&r=n^2\hbar^2/GMm^2.\end{aligned}\tag{41}$$

Deviations from this behavior, such as arise from gravitational spin-orbit coupling, are to be expected only when the calculated velocity of the electron in the lowest Bohr orbit is comparable to the speed of light:

$$GMm/\hbar\sim c,$$

or

$$Mm\sim\hbar c/G=(2.18\times10^{-5}\text{ g})^2,$$

or

$$M\sim5\times10^{19}\text{ g}.\tag{42}$$

For the analysis to apply in such an extreme case it would be necessary that this attracting mass should be confined within a distance of the order of the Schwarz-

[9] See, for example, H. A. Bethe, *Handbuch der Physik* (Julius Springer Verlag, Berlin, 1953, second edition), Vol. 24, Part 1.

[10] E. Schrödinger, Commentationes Pontif. Acad. Sci. 2, 321 (1938).

[11] See for example, L. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), second edition.

schild radius of this mass, or the Compton radius of the electron,

$$r_{\text{Schw}}=2GM/c^2=2\hbar/mc=7.7\times10^{-11}\text{ cm}$$

—a condition impossible of attainment even with matter of nuclear density.

The foregoing weak field analysis demands a binding energy for the particle small compared to its rest energy. It obviously will not apply to the neutrino with its zero rest mass. Moreover, bound orbits lie in the energy region between $+mc^2$ and $-mc^2$ and will cease to exist for an object with zero rest mass. The wave function for such an object never falls off exponentially in the region where the metric has become flat. However, it is possible to construct a metric with an inner region, a barrier region, and an outer region, in such a way that the neutrino wave function falls off exponentially in the barrier region. Then leakage from the inner region to the outer region is greatly inhibited. Effectively bound proper states of long life then exist for the neutrino in the inner region.

Such trapping of neutrinos is illustrated especially simply in the metric of the thin shell spherical geon,[12]

$$\begin{aligned} &e^{\nu}=e^{-\lambda}=1-(2GM/c^2r) \quad \text{for} \quad r>(9/4)(GM/c^2),\\ &e^{\nu}=1/9;\quad e^{-\lambda}=1 \quad \text{for} \quad r<(9/4)(GM/c^2). \end{aligned} \tag{43}$$

With the abbreviations

$$\begin{aligned} dr^*&=e^{\frac{1}{2}\lambda-\frac{1}{2}\nu}dr,\\ \rho&=(c^2/GM)r,\\ h\ (\text{cm}^{-1})&=\text{energy}/\hbar c,\\ \epsilon\ (\text{dimensionless})&=(GM/c^2)h\\ &=(GM/\hbar c^3)\ (\text{energy}), \end{aligned} \tag{44}$$

we rewrite the two first-order wave equations (39) for zero rest mass and zero electrostatic potential. We then eliminate one of the two dependent variables to obtain a single equation of the second order for the other variable; either

$$d^2F/d\rho^{*2}+[\epsilon^2-\xi(\rho)]F=0,$$

with

$$\xi(\rho)=e^{\nu}k^2/\rho^2-e^{\nu-\frac{1}{2}\lambda}k/\rho^2+(k/\rho)(d/d\rho^*)e^{\frac{1}{2}\nu}; \tag{45}$$

or

$$d^2G/d\rho^{*2}+[\epsilon^2-\eta(\rho)]G=0,$$

with

$$\eta(\rho)=e^{\nu}k^2/\rho^2+e^{\nu-\frac{1}{2}\lambda}k/\rho^2-(k/\rho)(d/d\rho^*)e^{\frac{1}{2}\nu}. \tag{46}$$

The last term on the right-hand sides of the dimensionless effective potentials $\xi(\rho)$ and $\eta(\rho)$ of (45) and (46) has the character of a spin-orbit coupling. As in the case of an electron moving under electrostatic forces, where the spin-orbit coupling is proportional to the angular momentum and to the radial derivative of the potential, so here one term in the effective potential experienced by the neutrino is proportional to the angular momentum parameter, k, and also proportional to the radial derivative of the metric quantity, $e^{\frac{1}{2}\nu}=(-g_{44})^{\frac{1}{2}}$. However, this term appears with opposite sign in the second order wave equations for the two components, F and G, of the *same* state function. From the fact that the two different equations have the same eigenvalues, it follows that the last *two* coupling terms together in (45) and in (46) produce no net effect on the energy levels. Otherwise stated, the energy eigenvalues are completely invariant against the change from k to $-k$. This degeneracy is the same as the fundamental polarization degeneracy that was discussed in Sec. 4. The demand that the neutrino have right-handed polarization means that the allowed state function is given neither by the solution of (45) and (46) for positive k, nor by the solution for negative k, but by the proper linear combination of these two solutions. However, (45) and (46) give the correct pieces out of which to construct the allowed total wave function, and also gives the correct energy eigenvalues.

It is now appropriate to check that the same radial equations are obtained in the two component formalism. For spin matrices, s_i, that satisfy the equations of definition (23), we make a choice analogous to the choice (1) of (28):

$$\begin{aligned} s_1&=e^{\frac{1}{2}\lambda}\begin{pmatrix}0&1\\1&0\end{pmatrix}=e^{\frac{1}{2}\lambda}\tilde{s}_1;\\ s_2&=r\begin{pmatrix}i&0\\0&i\end{pmatrix}=r\tilde{s}_2;\\ s_3&=r\sin\theta\begin{pmatrix}-1&0\\0&1\end{pmatrix}=r\sin\theta\tilde{s}_3;\\ s_4&=e^{\frac{1}{2}\nu}\begin{pmatrix}0&1\\-1&0\end{pmatrix}=e^{\frac{1}{2}\nu}\tilde{s}_4. \end{aligned} \tag{47}$$

To define covariant differentiation we need 2×2 matrices, Γ_k, that satisfy the analog of (2)

$$\partial s^i/\partial x^k-\Gamma_{ik}{}^{\mu}s_{\mu}-\Gamma_{k\ \text{comp. conj.}}s_i+s_i\Gamma_k=0, \tag{48}$$

namely,

$$\begin{aligned} &\Gamma_1=0;\quad \Gamma_2=\tfrac{1}{2}e^{-\frac{1}{2}\lambda}\tilde{s}_{1\ \text{comp. conj.}}\tilde{s}_2;\\ &\Gamma_3=\tfrac{1}{2}(\sin\theta e^{-\frac{1}{2}\lambda}\tilde{s}_{1\ \text{comp. conj.}}\tilde{s}_3+\cos\theta\tilde{s}_{2\ \text{comp. conj.}}\tilde{s}_3);\\ &\Gamma_4=(\nu'/4)e^{\frac{1}{2}\nu-\frac{1}{2}\lambda}\tilde{s}_{1\ \text{comp. conj.}}\tilde{s}_4. \end{aligned} \tag{49}$$

These representation-independent expressions evidently can also be generated from the Γ_k's, of the Dirac formalism (30) by a simple change: every product of the form $\tilde{\gamma}_i\tilde{\gamma}_j$ is replaced by a product of the form $\tilde{s}_{i\ \text{comp. conj.}}\tilde{s}_j$. The Dirac equation for a particle of mass m in the present formalism operates on a pair of spinors, ψ and φ, each of two components,

$$\begin{aligned} s^{\alpha}\nabla_{\alpha}\psi+\mu\varphi&=0,\\ s^{\alpha}{}_{\text{comp. conj.}}\nabla_{\alpha}\varphi+\mu\psi&=0, \end{aligned} \tag{50}$$

[12] J. A. Wheeler, Phys. Rev. 97, 511 (1955).

but the Pauli-Lee-Yang neutrino equation has the simpler form

$$0=s^\alpha\nabla_\alpha\psi=s^\alpha(\partial/\partial x^\alpha-\Gamma_\alpha)\psi. \quad (51)$$

This equation is obtained in explicit form from (31) by annulling the mass term, $\mu\psi$ and replacing $\tilde{\gamma}_i$ everywhere by $\tilde{s}_i$. Again introduce a new form for the wave function,

$$\omega=\exp(\nu/4)r(\sin\theta)^{\frac{1}{2}}\psi, \quad (52)$$

as in (33). Also multiply the equation for ω through by $\tilde{s}_{4\text{ comp. conj.}}$ and note that

$$\tilde{s}_{4\text{ comp. conj.}}\tilde{s}_4=-1,$$

and recall that the quantities

$$\tilde{s}_{4\text{ comp. conj.}}\tilde{s}_i=\sigma_i, \quad (i=1,2,3) \quad (53)$$

are the Pauli spin matrices. Then the general relativity form of the Pauli-Lee-Yang neutrino equation in a centrally symmetric metric becomes

$$(\sigma_1 e^{-\frac{1}{2}\lambda}\partial/\partial r+\sigma_2\partial/r\partial\theta +\sigma_3\partial/r\sin\theta\partial\varphi+e^{-\frac{1}{2}\nu}\partial/\partial T)\omega=0. \quad (54)$$

We consider a solution, ω, which depends upon azimuthal angle and time as

$$\exp(im_j\varphi-ihT),$$

so that (54) reduces to

$$(\sigma_1 e^{-\frac{1}{2}\lambda}\partial/\partial r+\sigma_2\partial/r\partial\theta+im_j\sigma_3/r\sin\theta-ihe^{-\frac{1}{2}\nu})\omega=0. \quad (55)$$

We can no longer find an angular operator, K, that commutes with (55). Instead, we proceed as follows. (1) We *temporarily* introduce an explicit representation for the Pauli matrices, σ_i, and express (55) as two coupled first-order equations for two unknown functions. (2) We deduce the character of the solution from what we already know of the solution in the Dirac formalism. (3) Out of this solution we construct the mixed density matrix, $M(1,2)=\psi(2)\psi^\dagger(1)$. From that matrix we can calculate the expectation value of any physical quantity. The neutrino flux, for example, is given by the trace

$$(\text{neutrino flux})^k=\psi^\dagger is^k\psi =Tr[is^kM(1,1)]. \quad (56)$$

(4) This density matrix can be expressed as a linear combination of the four matrices, σ_i and $\mathbf{1}$, and thus translated back to a form independent of any special representation of the Pauli matrices. It then takes the form

$$\begin{aligned}M(1,2)&=\psi(2)\psi^\dagger(1)\\ &=\exp[-\tfrac{1}{4}\nu(r_1)-\tfrac{1}{4}\nu(r_2)](r_1r_2)^{-1}\\ &\quad\cdot\{[\bar{F}(r_1)F(r_2)+\bar{G}(r_1)G(r_2)]\\ &\quad\times[\bar{\Theta}_1(1)\Theta_1(2)+\bar{\Theta}_2(1)\Theta_2(2)]\\ &\quad+i[\bar{G}(r_1)F(r_2)-\bar{F}(r_1)G(r_2)]\\ &\quad\times[\bar{\Theta}_1(1)\Theta_1(2)-\bar{\Theta}_2(1)\Theta_2(2)]\}\mathbf{1}(is^4)+\text{etc.}\end{aligned} \quad (57)$$

where F and G are the radial functions already introduced. The angular factors in (57) have the form

$$\begin{aligned}\Theta_1(1)&=f(\theta_1)\exp(im_j\varphi_1)/(\sin\theta_1)^{\frac{1}{2}},\\ \Theta_2(1)&=g(\theta_1)\exp(im_j\varphi_1)/(\sin\theta_1)^{\frac{1}{2}}, \text{ etc.,}\end{aligned} \quad (58)$$

where f and g are the angular functions of Schrödinger.[10] Such quantities as the number, flux and stress and energy density of neutrinos follow in a straightforward way from the density matrix (57) according to the pattern of (56), based on two component spinors. Alternatively the same answers can be obtained from the Dirac formalism, with which we shall generally work.

In either 2 or 4 component formalisms the energy levels of a trapped neutrino are found by solution of an eigenvalue equation which can be taken to be (45). We assume the geon metric (43). Then the dimensionless effective potential, $\xi(\rho)$, has the form

$$\xi(\rho)=[1-(2/\rho)](k^2/\rho^2)-[1-(2/\rho)]^{1/2}(k/\rho)[1-(3/\rho)]$$

outside the radius $\rho=2.25$; and inside it becomes

$$\xi(\rho)=k(k-1)/9\rho^2. \quad (59)$$

This potential is sketched in Fig. 1 for several values of the positive integer, k. Semistable bound states occur only for values of k of the order of 10 and larger. For such values of k terms in k can be neglected relative to terms in k^2. Then in the outer region $\xi(\rho)$ attains its maximum value

$$\xi_{\max}(\rho_{\max})=k^2/27$$

at

$$\rho_{\max}=3. \quad (60)$$

Its minimum

$$\xi_{\min}(\rho_{\min})=4^2k^2/9^3$$

occurs at

$$\rho_{\min}=9/4. \quad (61)$$

The dimensionless vibration frequency or energy, ϵ, of the nth semistable bound state can be estimated from the JWKB approximation formula,

$$(n+\tfrac{1}{2})\pi=\int_a^b[\epsilon^2-\xi(\rho)]^{\frac{1}{2}}d\rho^*, \quad (62)$$

where a and b are the classical turning points of the "bound" state of highest energy. The integral exists only when ϵ lies between $\xi_{\min}^{\frac{1}{2}}=2k/27$ and $\xi_{\max}^{\frac{1}{2}}=k/3^{\frac{3}{2}}$ (the two dashed limits in the right-hand part of Fig. 1). The total number, N_k, of bound states of type k can be estimated by setting ϵ^2 equal to the maximum of $\xi(\rho)$ and calculating (62), with the result

$$N=\pi^{-1}k\left\{\int_{9/4}^{3}[(1/27)-(1/\rho^2)+(2/\rho^3)]^{\frac{1}{2}}[1-(2/\rho)]^{\frac{1}{2}}d\rho +\int_{\sqrt{3}}^{9/4}[(1/27)-(1/9\rho^2)]^{\frac{1}{2}}3d\rho-\tfrac{1}{2}\right. =0.15k-0.5. \quad (63)$$

For each value of k the spectrum of allowed values for the dimensionless energy parameter, ϵ, stretches from a minimum value a little above $\xi_{\min}{}^{\frac{1}{2}}$ (proportional to k) to a maximum value a little less than $\xi_{\max}{}^{\frac{1}{2}}$ (also proportional to k). In this interval the number of levels, N_k, is also approximately proportional to k. These qualitative features of the level scheme are shown in Fig. 1.

In the absence of a gravitational field the neutrino spectrum reduces to the natural analog of the spectrum of the free electron. For each wave number, (k_x,k_y,k_z), there is one state of right-handed circular polarization and of positive energy; and a second state of negative energy and left handed circular polarization. All states of negative energy are to be considered as filled. Absence of a neutrino from a negative energy is to be interpreted as the presence of an antineutrino. According to hole theory the momentum (or angular momentum) of the real physical antineutrino is the negative of the momentum (or angular momentum) of the missing negative energy neutrino. For that negative energy neutrino the momentum and spin angular momentum are *opposite* in direction, according to the equation, $H=c(\boldsymbol{\sigma}\cdot\mathbf{p})$, of Lee and Yang. Therefore, as they show, the momentum and spin angular momentum have *opposite* directions for the real physical antineutrino.

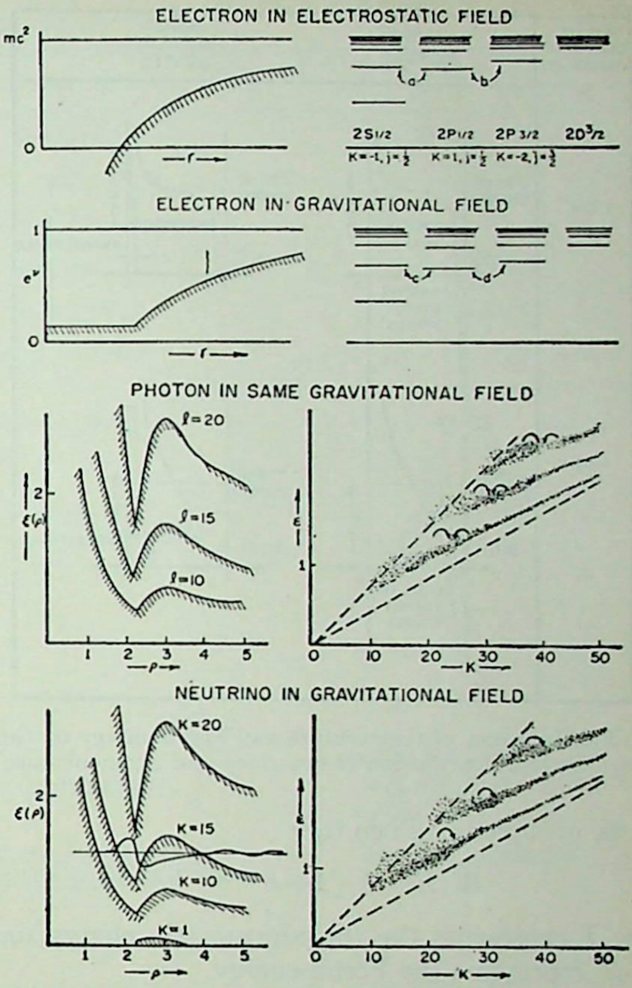


FIG. 1. Comparison of energy levels of an electron in an electrostatic field and in a gravitational field, and of a photon and a neutrino in a gravitational field. The left-hand diagram in these four cases gives, respectively, the electrostatic potential; the Einstein gravitational potential, $-g_{44}=e^\nu$ for the "geon metric"; the dimensionless effective potential, $\xi(\rho)=l(l+1)e^\nu/\rho^2$ for photons; and the corresponding quantity for neutrinos [the $\xi(\rho)$ of (59)]. The bound states of the electron lie lower than mc^2 and are therefore stable against electron escape. The energy splitting at (a) arises from Lamb shift and any other perturbative effects that raise the effective potential near the origin compared to the ideal Coulomb potential. Similarly, a splitting arises at (c) because s-wave electrons respond more than p-wave electrons to the central flat region of the metric. The splittings at (b) and (d) arise from spin-orbit coupling. Photons and neutrinos can be trapped only in states of greater or lesser life time, never in completely stable levels. The wave function of a trapped state falls off exponentially in the region of the effective potential barrier, lower left. Zero or small angular momentum corresponds to motion along, or nearly along, the radius vector—a kind of motion that always leads to escape of the photon or neutrino. The diagrams lower right show schematically both the position and the width of the semistable bound states. Each photon state can be occupied by any number of photons of either independent state of polarization. Each neutrino state can be occupied by only one neutrino. The corresponding antineutrino state has the same frequency and energy and can also be occupied. There is nothing to compare with the spin-orbit splitting of the electron states: the angular momentum of the neutrino always precesses in such a way as to stay parallel to the neutrino momentum.

6. STATISTICAL MECHANICS AND THERMAL EQUILIBRIUM OF NEUTRINOS

Gamow and Schoenberg[13] have given reason to believe that neutrino emission determines the rate of gravitational contraction of a heavy star in late phases of evolution, after the normal sources of thermonuclear energy have been exhausted. In oversimplified terms, hot neutrons change to cooler protons plus electrons plus neutrinos; hot protons and electrons change to cooler neutrons and antineutrinos. The medium continually loses energy by emission of neutrinos and antineutrinos—Gamow's "Urca process." The ν's and $\bar{\nu}$'s escape so much more readily than photons that they alone determine the rate of energy liberation and gravitational contraction. In late phases of such gravitational contraction the density might rise to a point where the opacity of matter even to neutrinos begins to make itself felt. For example, consider the point where nuclear densities have been reached, of the order of 10^{38} nucleons/cm^3; and assume a neutrino interception cross section, 10^{-43} cm^2, of the order of that found by Reines and Cowan[14]; then the mean free path of a neutrino will be of the order of only 1 km. Under such conditions one has to speak of an opacity with respect to neutrinos and a local neutrino temperature along lines familiar from the theory of transfer of heat by electromagnetic radiation. One does not have to consider this particular problem of stellar interiors to raise the question: what is the equilibrium distribution of neutrinos? The specification of this equilibrium demands one more quantity for neutrinos than for photons: one has to give both the temperature and a suitably defined Fermi energy. The probability $W(E)$ that a neutrino state of energy, E,

[13] G. Gamow and M. Schoenberg, Phys. Rev. 59, 539 (1941).
[14] F. Reines and C. Cowan, Jr., Phys. Rev. 92, 830 (1953).

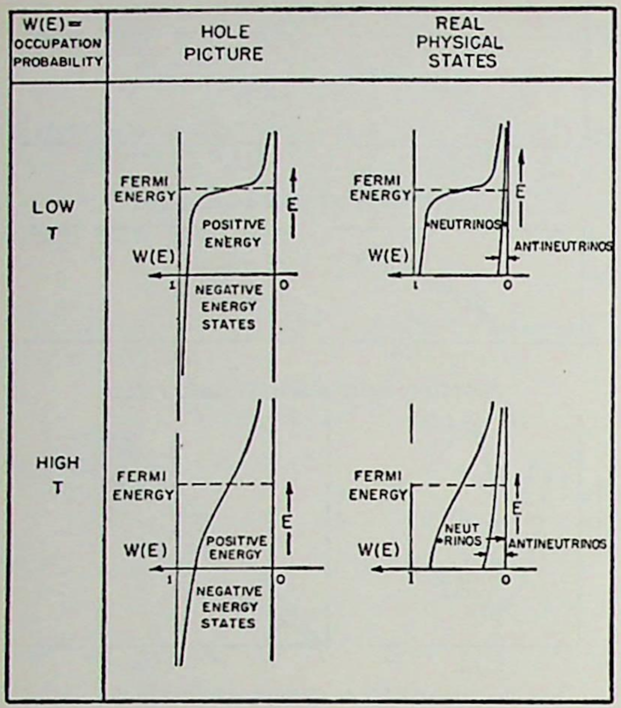


FIG. 2. Influence of temperature and Fermi energy on the equilibrium distribution of neutrinos and antineutrinos.

shall be occupied is given by[15]

$$W(E)=1/[1+e^{-\eta+E/T}], \tag{64}$$

where T represents the temperature (in energy units) and ηT represents the Fermi energy.

Translating from hole theory to the physically observable neutrino and antineutrino states (Fig. 2), one has the expression,

$$dn_+=[4\pi k^2dk/(2\pi)^3]/[1+e^{-\eta+k\hbar c/T}], \tag{65}$$

for the average number of neutrinos per unit volume in the interval, dk, of circular wave number; and for the corresponding number of antineutrinos,

$$\begin{aligned} dn_- &= [4\pi k^2dk/(2\pi)^3][1-W(-k\hbar c)] \\ &= [4\pi k^2dk/(2\pi)^3]/[1+e^{\eta+k\hbar c/T}]. \end{aligned} \tag{66}$$

The quantities of greatest interest are the total density of neutrinos,

$$n_+=(T^3/2\pi^2\hbar^3c^3)\int_0^\infty \frac{x^2dx}{1+e^{x-\eta}}, \tag{67}$$

the total density of antineutrinos (given by the same expression with the sign of η reversed), and the total energy density,

$$\epsilon=(T^4/2\pi^2\hbar^3c^3)\left[\int_0^\infty \frac{x^3dx}{1+e^{x-\eta}}+\int_0^\infty \frac{x^3dx}{1+e^{x+\eta}}\right]. \tag{68}$$

If the Fermi energy is zero, the integrals have the simple values

$$n_+=n_-=(T^3/2\pi^2\hbar^3c^3)\times 1.803,$$

and

$$\begin{aligned} \epsilon_\nu &= (T^4/2\pi^2\hbar^3c^3)2\times 5.757 \\ &= (7\pi^2/120)(T^4/\hbar^3c^3). \end{aligned} \tag{69}$$

In contrast, the energy content of electromagnetic blackbody radiation is

$$\epsilon_{em}=(8\pi^2/120)(T^4/\hbar^3c^3). \tag{70}$$

The neutrino energy is smaller than the electromagnetic energy by the factor $\frac{7}{8}$. The number of accessible states of a given wave number is the same for the two kinds of radiation: two spin polarizations for photons; and for the more penetrating radiation, one neutrino and one antineutrino state. Moreover, the occupation probability for either kind of state follows the same limiting Boltzmann formula for states of high energy $W(E)\sim e^{-E/T}$. But any given state of low energy can be occupied by many photons, and at most one neutrino: hence the advantage for electromagnetic energy compared to neutrino energy. These relationships are entirely changed when the Fermi energy of the neutrinos has a nonzero value. In that case the energy density of the neutrinos can be made to have any arbitrarily large value.

7. CREATION OF NEUTRINO PAIRS

In this paper we disregard all processes for creation or disappearance of neutrinos which depend upon transformations of the elementary particles, considering only the response of neutrinos to the curvature of the metric. More specifically, we limit attention to processes in which—according to hole theory—a neutrino is raised from a negative energy state to a positive energy state, or where, in physical terms, a neutrino-antineutrino pair is created.

Is there anything special about the gravitational field which makes it incapable of raising a neutrino from a negative energy state to a positive energy state? Otherwise stated, does there exist any quantity, Q, which will commute or anticommute with the operator in the wave equation (51) and which will serve to distinguish positive and negative energy states? To construct such an operator one has only the four independent 2×2 spin matrices. Among these only the unit matrix commutes with (51). From it nothing of interest can be constructed. We therefore expect that there is no point of principle which prevents transitions from negative energy to positive energy states. In other words, it is as much out of the question in neutrino physics as it is in electron physics to make a well-defined distinction between negative and positive energy states when general gravitational or electromagnetic fields are at work.

The mechanism of the transitions in the neutrino case is simple. Just as a static disturbance in the metric deflects a neutrino, bringing about a transition between states of different momenta, so a time-varying disturbance causes a transition between states of different

[15] E. Fermi, Z. Physik 36, 902 (1926).

TABLE III. Order-of-magnitude estimate of cross section for creation of a pair by the collision of two quanta of equal but opposite momenta.

Process	$\gamma+\gamma\rightarrow e^{+}+e^{-}$	$G+G\rightarrow\nu+\bar{\nu}$
Energy of one quantum	$E=\hbar c/\lambdabar$	$E=\hbar c/\lambdabar$
Localization volume	$\sim\lambdabar^3$	$\sim\lambdabar^3$
Energy density	$\sim\hbar c/\lambdabar^4$	$\sim\hbar c/\lambdabar^4$
Relevant field	electric	gravitational
Square of field	$\sim(\hbar c)^{\frac{1}{2}}/\lambdabar^2$	$\sim(\hbar cG)^{\frac{1}{2}}/\lambdabar^2$
Available potential in region of energy concentration	$\delta A\sim(\hbar c)^{\frac{1}{2}}/\lambdabar$	$\delta g\sim(\hbar G/c^3)^{\frac{1}{2}}/\lambdabar$
Potential required to produce transition from $E_1=-\hbar c/\lambdabar^2$ to $E_2=+\hbar c/\lambdabar^2$ with nearly 100% probability	$\delta A\sim\hbar c/\lambdabar e$	$\delta g\sim1$
Available disturbance/required disturbance	$\sim e/(\hbar c)^{\frac{1}{2}}$	$\sim(\hbar G/c^3)^{\frac{1}{2}}/\lambdabar$
Number of times this factor enters into matrix element	2	2
Number of times matrix elements occur in transition probability	2	2
Pair creation cross section for 100% creation probability	$\sim\lambdabar^2$	$\sim\lambdabar^2$
Resulting estimate for cross section for pair creation	$\sigma\sim(e^2/\hbar c)^2\lambdabar^2$ $\sim(e^2/mc^2)^2(mc^2/E)^2$	$\sigma\sim(\hbar G/c^3)^2/\lambdabar^2$ $\sim(1.6\times10^{-33}\text{ cm})^4/\lambdabar^2$ $\sim(GE/c^4)^2$
Asymptotic behavior of accurate formula for cross section, high energy, unpolarized radiation	$\sigma\sim2\pi(e^2/mc^2)^2(mc^2/E)^2$ $\times\ln[2E/e^{\frac{1}{2}}mc^2]$	Not yet calculated

energies. Let such a time-varying disturbance in the free gravitational field be analyzed into Fourier components in the weak field approximation. According to the laws of conservation of momentum and energy, no single monochromatic disturbance will be able to produce transitions of neutrinos from negative energy states to positive energy states (apart from the singular case where all three momenta lie along the same line). Real pair production demands collaboration of at least two of these monochromatic disturbances, traveling in different directions. As in the case where two photons collide to produce an electron pair,[16] so here it is simplest to analyze the two quantum process in a frame of reference in which the momenta are equal and opposite. For an order-of-magnitude estimate of the cross section for pair creation it is not necessary to evaluate matrix elements in detail. Instead, one can follow the reasoning sketched by Bohr and Rosenfeld in their discussion of vacuum polarization.[17] Let the two concentrations of energy be pictured as localized in two regions of space, of dimensions $\sim\lambdabar$, and as moving towards each other with velocity c, so that they overlap and collaborate only during a time interval $\sim\lambdabar/c$. Then the calculation outlined in Table III leads to a cross section of the order

$$\sigma\sim(GE/c^4)^2 \sim(1.6\times10^{-33}\text{ cm})^4/\lambdabar^2, \tag{71}$$

for creation of a $(\nu,\bar{\nu})$ pair by two gravitons, each of energy E. The cross section is fantastically small for quanta of any familiar energies. As the energy is increased, a domain of wavelengths is ultimately reached, $\lambdabar\sim(\hbar G/c^3)\sim1.6\times10^{-33}$ cm, at which any normal analysis of gravitational disturbances into waves would appear to be ruled out: the disturbances in the metric have reached the order of magnitude, $\delta g\sim1$, where nonlinear effects completely dominate the analysis. Even at such incredibly high energies the estimated cross section (71) only attains a value of the order

$$\sigma\sim10^{-66}\text{ cm}^2. \tag{72}$$

The principle of microscopic reversibility gives a cross section also of the order (71) for the process

$$\nu+\bar{\nu}\rightarrow G+G. \tag{73}$$

The existence of the inelastic process (73) implies, according to dispersion theory,[18] the existence of an elastic scattering process

$$\nu+\bar{\nu}(\text{via}\rightarrow G+G)\rightarrow\nu'+\bar{\nu}'. \tag{74}$$

[16] G. Breit and J. A. Wheeler, Phys. Rev. 46, 1087 (1934). See E. J. Williams, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. 13, 44 (1935) for the observable physical consequences of this pair creation process.

[17] N. Bohr and L. Rosenfeld, Phys. Rev. 78, 794 (1950).

[18] J. S. Toll, "The dispersion relation for light and its application to problems involving electron pairs," Ph.D. thesis, Princeton, 1952, unpublished; see also J. Toll and J. A. Wheeler, Phys. Rev. 81, 654(A) (1951).

To estimate the cross section for this process even for neutrino energies, $\hbar\omega$, small compared to the critical energy $(\hbar c^5/G)^{\frac{1}{2}}$, it is not legitimate to apply

$$(d\sigma/d\Omega)_{\text{elastic forward at }\omega}=\left[\frac{\omega^2}{2\pi^2 c}\int_0^\infty \frac{\sigma_{\text{abs}}(\omega')d\omega'}{\omega'^2-\omega^2}\right]^2, \quad (75)$$

because the principal contribution to the integral comes from virtual processes at energies, $\hbar\omega'$, comparable to and larger than the critical energy.

We have not considered processes that belong in the realm of true elementary particle physics; processes such as

$$\nu+\bar{\nu}\rightarrow\mu^+ + e^-, \quad (76)$$

and

$$\nu+\bar{\nu}(\text{via virtual}\rightarrow\mu^+ + e^-)\rightarrow\nu'+\bar{\nu}'. \quad (77)$$

In the frame where the total momentum is zero the first process has a threshold at a neutrino energy,

$$E=104mc^2. \quad (78)$$

The calculated cross section rises well above threshold according to the asymptotic formula[12]

$$\sigma_{\text{abs}}\sim(g^2/\hbar^2c^4)(E/mc^2)^2, \quad (79)$$

where the beta-coupling constant has the familiar value

$$g\sim 10^{-49}\text{ erg cm}^3. \quad (80)$$

From the absorption process one infers via causality arguments the existence of the scattering process (77). Applied to this process, the dispersion integral (76) is only logarithmically divergent. Attribute to the divergent logarithm the conventional value "10." Then one estimates[12] a scattering cross section of the order

$$(d\sigma/d\Omega)_{\text{forward elastic}}\sim\text{"100"}(E^3g^2/\hbar^5c^5)^2. \quad (81)$$

It appears reasonable to conclude that both $(\nu,\bar{\nu})$ scattering processes are negligible at any reasonable energies in comparison with the two absorption processes. The two absorption cross sections have the same E^2 energy dependence, interestingly enough. Their ratio is the dimensionless number

$$\frac{\sigma(\nu+\bar{\nu}\rightarrow\mu^+ + e^-)}{\sigma(\nu+\bar{\nu}\rightarrow G+G)}\sim\frac{(g^2/\hbar^2c^4)(E/\hbar)^2}{(GE/c^4)^2}$$
$$=(gc^2/G\hbar^2)^2\sim 10^{34}, \quad (82)$$

a testimonial to the well-known great ratio between beta couplings and gravitational couplings.

8. DENSITY OF STRESS, MOMENTUM, AND ENERGY

So far we have considered the response of a neutrino (or electron) to a given metric, either static or varying in time. Now we ask for the response of the gravitational field to the neutrino (or electron). The Einstein field equations connect the gravitational field and the change in the metric with the stress-momentum-energy tensor, T_{ik}. This tensor has been given for example by Pauli[19] in the case of flat space. Rosenfeld[20] has discussed this tensor in curved space from the viewpoint of the Vierbein formalism and derived some of its components. We follow his very general approach. The Lagrangian density in general relativity is proportional to the quantity

$$[(c^4/G)R+L](-g)^{-\frac{1}{2}}, \quad (83)$$

where R is the curvature scalar and L arises from other field variables in addition to the g_{ik}. The requirement that the Lagrangian be an extremum with respect to variations of the g_{ik} leads to the Einstein field equations of the form

$$G_{ik}(\equiv R_{ik}-\tfrac{1}{2}g_{ik}R)=\frac{8\pi G}{c^4}T_{ik}, \quad (84)$$

where the stress energy tensor, T_{ik}, is defined for ordinary fields as the variational derivative

$$T_{ik}=(1/8\pi)(-g)^{\frac{1}{2}}\frac{\delta(-g)^{-\frac{1}{2}}L}{\delta g^{ik}}$$
$$=(1/8\pi)\left(\frac{\delta L}{\delta g^{ik}}-\tfrac{1}{2}g_{ik}L\right), \quad (85)$$

and more generally, for spinor fields, by the equation

$$\int T_{\mu\nu}\delta g^{\mu\nu}(-g)^{\frac{1}{2}}d^4x=(1/8\pi)\delta\int L(-g)^{\frac{1}{2}}d^4x. \quad (86)$$

The procedure for finding this variation for spinor fields is somewhat more subtle than for c-number fields. The variation of the g_{ik} reflects itself not only where they are used to form invariants, but also in variations of the spinors which are necessary to preserve the fundamental relationship (1): we must have

$$[\delta\gamma_i,\gamma_j]_+ + [\gamma_i,\delta\gamma_j]_+ = 2\delta g_{ij}\mathbf{1}. \quad (87)$$

One solution is[21]

$$\delta\gamma^i=\tfrac{1}{2}\gamma_\alpha\delta g^{\alpha i}. \quad (88)$$

Since any other solution can be obtained from this one by a similarity transformation, the above is general enough for our purposes. This variation of the γ^i leads, through (2), to the variation of the Γ_k,

$$\delta\Gamma_k=\tfrac{1}{8}(g_{\nu\sigma}\delta\Gamma_{\mu k}{}^{\sigma}-g_{\mu\sigma}\delta\Gamma_{\nu\rho}{}^{\sigma})s^{\mu\nu}, \quad (89)$$

where s^{ij} is an abbreviation for $\frac{1}{2}(\gamma^i\gamma^j-\gamma^j\gamma^i)$. For the

[19] W. Pauli, *Handbuch der Physik* (Julius Springer Verlag, Berlin, 1933), second edition, Vol. 24, Part 1, p. 235.

[20] L. Rosenfeld, Acad. roy. Belg. 18, No. 6 (1940). We are indebted to Bryce DeWitt for the opportunity to see an unpublished manuscript of his written several years ago in which the result (92) has already been given.

[21] W. Pauli, Ann. Phys. 18, 337 (1933).

Lagrangian (7), which vanishes when the field equations hold, we find the variation

$$\delta\int L(-g)^{\frac{1}{2}}d^4x=\int \delta L(-g)^{\frac{1}{2}}d^4x$$

$$=\int[\psi^\dagger\delta\gamma^k\nabla_k\psi+\psi^\dagger\gamma^k\delta\Gamma_k\psi](-g)^{\frac{1}{2}}d^4x$$

$$=\int[\tfrac{1}{2}\psi^\dagger\gamma_\nu\nabla_\mu\psi\delta g^{\mu\nu}+\tfrac{1}{8}(g^{\rho\nu}s^\mu-g^{\rho\mu}s^\nu)$$

$$\times(g_{\mu\sigma}\delta\Gamma_{\nu\rho}{}^\sigma-g_{\nu\sigma}\delta\Gamma_{\mu\rho}{}^\sigma)](-g)^{\frac{1}{2}}d^4x, \quad (90)$$

where s^k is the previously defined four vector of flow and density. The $\delta\Gamma_{\nu\rho}{}^\sigma$ are expressible in terms of the $\delta g^{\mu\nu}$ and their derivatives. The derivatives are removed by partial integration, and the various terms in the second part of the integrand collected to form the covariant derivative of the typical component of the flow vector, s^k:

$$\delta\int L(-g)^{\frac{1}{2}}d^4x$$

$$=\int[\tfrac{1}{2}\psi^\dagger\gamma_\nu\nabla_\mu\psi-\tfrac{1}{4}\nabla_\mu s_\nu]\delta g^{\mu\nu}(-g)^{\frac{1}{2}}d^4x. \quad (91)$$

Comparing (91) with (86), we find the momentum-energy tensor

$$T_{ik}=-(\hbar c/4)[\psi^\dagger\gamma_i\nabla_k\psi-(\nabla_k\psi^\dagger)\gamma_i\psi$$
$$+\psi^\dagger\gamma_k\nabla_i\psi-(\nabla_i\psi^\dagger)\gamma_k\psi]. \quad (92)$$

In the case of special relativity, where $\nabla_k=\partial/\partial x^k$, this result—which holds both for the electron and the neutrino—reduces to the energy-momentum tensor given by Pauli.[19] The trace of the energy-momentum tensor (92) reduces to

$$T_\alpha{}^\alpha=\mu\hbar c\psi^\dagger\psi \quad (93)$$

as a consequence of the field equations. It is remarkable that this trace vanishes for the neutrino field ($\mu=0$) just as it does for the electromagnetic field.

One can use this result to derive the expressions already given for the density of neutrino energy under conditions of thermodynamic equilibrium.

9. INTERACTION BETWEEN TWO PENCILS ON NEUTRINOS

We consider a second application of the energy momentum tensor. We ask for the effect of a directed beam of neutrinos in an otherwise nearly flat space. For simplicity, let the beam move in the x direction. Take $k_1=-k_4=k^4=k=E/\hbar c$. Then the neutrino wave has the form

$$\psi=u\exp[ik(x^1-x^4)]. \quad (94)$$

Here u is a slowly varying spinor function of position which (1) is nearly constant in the domain of the pencil of radiation—very large compared to $\lambda\!\!\!^-=1/k$—and (2) falls off smoothly and strongly outside of this region. The gradient of ψ contains terms that come from the gradient of u, and others from the gradient of the exponential; but the second terms are overwhelmingly more important than the first ones. We conclude that all components of T_{ik} with i or k equal to 2 or 3 are negligible. The significant components have the value

$$T_{14}=T_{44}=T_{11}=\hbar ck|\psi|^2, \quad (95)$$

since the flow-density four vector is light-like. Tolman and Ehrenfest have investigated[22] the metric due to a pencil like concentration of electromagnetic energy of identical character. Therefore the results which they derived for light also hold for neutrinos. Two photons or neutrinos, or one electron and one neutrino, attract when their propagation vectors are antiparallel with twice the Newtonian value, and not at all when their propagation vectors are parallel. Therefore the neutrinos in a toroidal neutrino geon will be in their most stable configuration when half go around one way and half go around the other way.

10. STRESS, MOMENTUM, AND ENERGY OF A NEUTRINO IN A TRAPPED STATE

As a third application of the stress energy tensor we consider the motion of a neutrino in a nearly bound orbit in a central field, as in Sec. 5. Let the time dependence of the wave function be described by the factor $\exp(-iET/\hbar c)$, and neglect any small imaginary part of E that describes the slow leakage of the state out of the zone of trapping. Then the stress energy tensor takes the form

$$\begin{aligned} T_r{}^r&=(i\hbar c/2)e^{-\frac{1}{2}\lambda}[(\partial\psi^*/\partial r)\tilde{\gamma}_4\tilde{\gamma}_1\psi \\ &\qquad -\psi^*\tilde{\gamma}_4\tilde{\gamma}_1(\partial\psi/\partial r)], \\ T_T{}^T&=-e^{-\frac{1}{2}\nu}E\psi^*\psi, \\ T_\theta{}^\theta+T_\varphi{}^\varphi&=-[T_r{}^r+T_T{}^T]. \end{aligned} \quad (96)$$

Require that ψ be an eigenstate of the z component of the total angular momentum. Then these expressions are independent of ϕ. It is therefore a simple matter to sum or average them over various orientations of the orbit provided the contributions add incoherently. It is easy to give conditions which make this incoherent addition legitimate.[12] The effect of the addition or averaging is to make $\langle T_r{}^r\rangle$, $\langle T_T{}^T\rangle$, $\langle T_\theta{}^\theta\rangle$, and $\langle T_\varphi{}^\varphi\rangle$ the only nonzero components, and to eliminate the dependence upon angle. Then these surviving components of the averaged stress-momentum-energy tensor depend only upon the radial components of the neutrino wave function. When this wave function has the form (36), using the wave equation (32), these averages take

[22] See, for example, R. C. Tolman, ***Relativity, Thermodynamics and Cosmology*** (Oxford University Press, New York, 1934), p. 272.

the form

$$\langle T_r{}^r\rangle = (e^{-\frac{1}{2}\nu}/r^2)[e^{-\frac{1}{2}\nu}ER^*R-(\hbar k/r)R^*\tilde{\gamma}_1 R]$$
$$\langle T_T{}^T\rangle = -(e^{-\nu}/r^2)ER^*R, \quad (97)$$
$$\langle T_\theta{}^\theta\rangle = \langle T_\varphi{}^\varphi\rangle = (\hbar k/r^2)e^{-\frac{1}{2}\nu}R^*\tilde{\gamma}_1 R;$$

however, this form applies only to a Dirac neutrino. The normalization of (36) is such that the fourth component of the flow-density vector integrates to unity:

$$1=\int s^4 d(\text{volume})$$
$$=-\int[\langle T_T{}^T\rangle/E]d(\text{volume})$$
$$=\int_0^\infty R^*Re^{\frac{1}{2}\lambda-\frac{1}{2}\nu}dr\int\frac{\Theta^*\Theta}{\sin\theta}\sin\theta d\theta d\varphi. \quad (98)$$

The radial and angular integrals are separately normalized to unity here and hereafter.

The wave function for the right-handed neutrino, as already noted, does not have the form, ω, of (36), but rather, the form, $2^{-\frac{1}{2}}(1-i\gamma_5)\omega$ of a linear combination of two parts. Each part is individually separable in the product form but their sum is not. Let R continue to denote the radial factor in the first of these two parts. Then the normalization (98) will continue to apply. The form (97) will not be valid for the stress-energy tensor for any individual neutrino state; new interference terms have to be added. However, when the average is made, as before, over orbits of all orientations, and use is made of the angular properties of the functions, as deduced by Schrödinger, then these interference terms drop out, and (97) continues to apply to the Lee-Yang neutrino.

11. NEUTRINO GEON

So far as one knows, geons have nothing to do either with elementary particle physics or with astrophysics. They have interest for quite another reason. Gravitational fields have a nonlinear character that has many unexplored consequences. A few of these consequences come to light in the analysis of a geon.[23] A collection of immaterial energy holds itself together by its own gravitational attraction. The analysis of such an object or "geon" combines both major topics of the present paper: the response of neutrinos to gravitational fields, and the production of gravitational fields by neutrinos.

Geons of toroidal form appear in principle to be more stable than spherical geons, at least when comparable numbers of quanta go around in the two opposite directions. The case in which the angular momentum is zero has been analyzed by Ernst.[24]

In the opposite extreme case where all the neutrinos go around the same way, then the right-handed character of the neutrinos implies that the resulting toroidal system is not mirror symmetric with respect to reflections in its own plane. It would be interesting to investigate the consequences of this asymmetry to see whether it shows up in the gravitational field near the toroidal geon.

Among spherically symmetrical geons perhaps the simplest to consider is a thermal geon.[23] A thermal neutrino geon differs from a thermal electromagnetic geon only in a trivial respect. The photon energy density in a thermal geon is

$$T_T{}^T=(\pi^2/15)(T^4/\hbar^3c^3)e^{-2\nu}(\omega/4\pi), \quad (99)$$

where T is the temperature (in energy units), $-e^{-\nu(r)}=g_{44}$ is the familiar factor in the metric, and $\omega/4\pi$ is the solid angle within which those rays travel which are trapped. The neutrino energy density in a thermal neutrino geon is similarly

$$T_T{}^T=(\pi^2/15)(T^4/\hbar^3c^3)e^{-2\nu}(\omega/4\pi)f(\eta). \quad (100)$$

Here the dimensionless factor,

$$f(\eta)=(15/2\pi^4)\left[\int_0^\infty\frac{x^3dx}{1+e^{x-\eta}}+\int_0^\infty\frac{x^3dx}{1+e^{x+\eta}}\right], \quad (101)$$

is uniquely fixed by the Fermi energy, $E_F=\eta T$ of the trapped neutrinos. All the analysis of thermal electromagnetic geons[23] can be taken over immediately to thermal neutrino geons. For this purpose it is only necessary to redefine the characteristic scale of length, R_T, as

$$R_T=(15\hbar^3c^7/8\pi^3 f(\eta)GT^4)^{\frac{1}{2}}, \quad (102)$$

and the characteristic unit of mass as

$$M_T=(15\hbar^3c^{11}/8\pi^3 f(\eta)G^3T^4)^{\frac{1}{2}}. \quad (103)$$

Then all the calculated graphs of reference 23 apply at once, and a complete description of the metric and energy distribution is available for a thermal neutrino geon.

In the other simple limiting case of a spherical neutrino geon, all of the energy is concentrated in a thin spherical shell. In ray language, all of the neutrinos are executing circular orbits of the same radius. Orbits of all orientations occur with equal probability. It is a matter of choice—that is, a matter of initial conditions—whether neutrino orbits of many different energies are occupied, or only states of a single energy are filled. We take the second option to simplify the analysis, although it demands a higher energy per neutrino to provide a given total mass. The wavelength of the neutrinos is then very small compared to the geon radius so the quantum number, k, is very large.

To formulate equations for the self-consistent metric, we proceed as in reference 12, introducing the ab-

[23] For further discussion, especially of unexplored analogies between gravitation physics and hydrodynamics, see E. Power and J. A. Wheeler, Revs. Modern Phys. 29, 480 (1957).

[24] F. Ernst, Revs. Modern Phys. 29, 496 (1957).

breviations

$$\Omega = E/\hbar c,$$
$$\rho = \Omega r, \quad (104)$$
$$d\rho^* = e^{\frac{1}{2}\lambda - \frac{1}{2}\nu} d\rho,$$
$$e^{-\lambda} = 1 - 2\rho^{-1} L(\rho),$$

where $L(\rho)$ is a parameter that measures the mass included within the distance r;

$$e^{\lambda+\nu} = Q^2(\rho);$$

and

$$\begin{pmatrix} f(\rho) \\ g(\rho) \end{pmatrix} = [8\pi G(2k-1)E/c^4]^{\frac{1}{2}} \begin{pmatrix} F(r) \\ G(r) \end{pmatrix}, \quad (105)$$

where the radial functions F and G have the normalization

$$\int_0^\infty e^{\frac{1}{2}\lambda - \frac{1}{2}\nu} (F^2 + G^2) dr = 1. \quad (106)$$

The wave equation takes the form

$$d^2 f/d\rho^{*2} + [1 - (e^\nu k^2/\rho^2) + (d/d\rho^*)(e^{\frac{1}{2}\nu} k/\rho)] f = 0, \quad (107)$$

and the Einstein field equations lead to the formulas

$$dL/d\rho^* = (1/2Q)(f^2 + g^2)$$
$$= (1/2Q)[f^2 + (df/d\rho^*)^2 + (e^\nu k^2/\rho^2) f^2 - (e^{\frac{1}{2}\nu} k/\rho) df/d\rho^*],$$
$$dQ/d\rho^* = (\rho - 2L)^{-1} [f^2 + (df/d\rho^*)^2]. \quad (108)$$

These equations for the self-consistent field of a thermal geon are very similar to (38), (40), and (41) of reference 12 for electromagnetic geons. When we neglect terms in k compared to terms in k^2, the two systems of equations become identical. The solution for the case of large angular momentum[12] leads to the geon metric already used (43) in discussing the response of neutrinos to a given gravitational field. In summary, geons can be constructed in principle out of neutrinos in much the same way that they can be constructed in principle out of photons.

12. CONCLUSION

From our analysis of some of the interactions between neutrinos and gravitational fields we conclude that neutrino physics has an interesting character in and by itself, even when attention is withdrawn from all beta-ray transformations. The behavior of neutrinos has become a little clearer, but the mystery why spinors occur in nature is left as pressing as ever. What is there about the description of the geometry of space which is not already adequately covered by ordinary scalars, vectors, and tensors of standard tensor analysis? To this question the mathematics of spinor fields gives a well known answer: spinors allow one to describe rotations at one point in space completely independently of rotations at all other points in space—rotations that have nothing to do with the coordinate transformations that are treated in the usual tensor analysis. Fully to see at work this machinery of independent rotations at each point in space, we do best to consider the spinor field in a general curved space, as in this paper. But the deeper part of such rotations in the description of nature is still mysterious.

13. ACKNOWLEDGMENTS

We wish to express our appreciation to Professor V. Bargmann for his interest and for several illuminating discussions; also to Professor C. N. Yang for discussion of the equivalence between the Dirac and the two-component spinor formalism.

Thermal Geons

EDWIN A. POWER, *Department of Mathematics, University College, London*

AND

JOHN A. WHEELER, *Palmer Physical Laboratory, Princeton University, Princeton, New Jersey*

1. INTRODUCTION AND SUMMARY

THE well-established classical theory of fields with zero rest mass has rich consequences, many of which still stand unexplored today (Table I), forty years after Einstein's formulation of the general relativity theory of electromagnetism and gravitation.

One new consequence appeared in a recent investigation.[1] An electromagnetic field, or a neutrino field, or a mixed field, of appropriate character and sufficient energy density can hold itself together, it was calculated, by its own gravitational attraction for a time long in comparison with the characteristic periods of the field oscillations. Or a the gravitation field, sufficiently strong, can guide an electromagnetic or neutrino wave and confine its energy to a bounded region of space. When the energy of the standing wave is great enough, it has enough mass to provide the guiding gravitational field all by itself. The wave holds itself together. Thus the gravitational and electromagnetic field equations of general relativity admit self-consistent solutions of great variety, many of which describe a reasonably stable concentration of energy. Such a "gravitational electromagnetic entity" or "geon" is endowed with mass, has a characteristic decay rate in the free state, moves through space like a Newtonian "body" when subjected to fields that vary sufficiently slowly in space and time, and under the influence of stronger fields undergoes transmutation. Geons of mass greater than 10^{38} g and radius over 10^{11} cm are subject to classical analysis. Smaller geons, even very much smaller geons, presumably exist, but are specifically quantum objects. Their properties have not yet been investigated. The large, classical geons have no known connection with observational science. Their interest lies in what they tell of the richness of a pure field theory. *The* pure classical field theory, it would perhaps be better to say: the only well tested theory that one has for fields of zero rest mass; the theory of elementary fields, not fields associated with particles that are themselves complex; the only field theory long established in its own right, not invented to account for selected particles or for special features of nuclear forces.

From the pure field theory of the fields of zero rest mass, and the geon concept, this article aims only at a detailed analysis of one particular kind of geon, a *thermal* geon, with the following properties:

(1) The gravitational field is static and spherically symmetric; i.e., in appropriate coordinates the element of proper distance, ds, or proper cotime, $d\tau$, has the form,

$$(ds)^2 = -(d\tau)^2 = g_{\alpha\beta}dx^\alpha dx^\beta = e^\lambda (dr)^2 + r^2[(d\theta)^2 + (\sin\theta d\phi)^2] - e^\nu (dT)^2, \quad (1)$$

where λ and ν are functions of r alone, found by numerical integration.

(2) Each independent mode of vibration of the electromagnetic field is idealized to fall into one or other of two sharply separated classes. The energy of modes of the first class remains trapped for all time. The second class of vibrations carry energy freely to infinity. The one class corresponds (in the language of photon orbits) to bounded null geodesics, the other to null geodesics that lead to infinity (Fig. 1). In actuality the energy of the "bounded" modes leaks off to infinity at a nonzero rate through a refractive index barrier, after the manner of alpha-particle penetration through a nuclear potential barrier. The rate of leakage falls off exponentially as the ratio of the dimensions of the geon to the wavelength of the disturbance in question. In a thermal geon the wavelength of the average mode is so small in comparison with the geon's radius that the leakage rate is effectively zero for all but a very restricted class of vibration modes. These few modes on the orbit picture belong to geodescis close to the crossover from bounded orbits to free orbits. The outward transport of energy resulting from such modes is relatively small and is legitimately neglected in a first analysis of a thermal geon.

(3) All free modes are assigned zero energy. Each bound mode of circular frequency Ωc is idealized to have an excitation

$$m_\Omega c^2 = E_\Omega = \hbar c\Omega[-1+\exp(\hbar c\Omega/T)]^{-1}. \quad (2)$$

All properties of the thermal geon are specified by the single parameter T, which has the dimensions of energy and which we call the temperature. Occurrence of the quantum of angular momentum, $\hbar$, in this formula makes assignment of boundary conditions no different in principle from that in any classical problem: It is an act from outside, it may or may not have a quantum origin, but it alters in no way the purely classical character of all the *rest* of the analysis.

It is easy to compare and contrast thermal geons with other gravitational-electromagnetic entities. In all such objects effective confinement of the energy

[1] J. A. Wheeler, Phys. Rev. 97, 511 (1955).

TABLE I. Qualitative comparison of the present state of charge- and mass-free gravitation and electromagnetism with hydrodynamic theory showing how many features of the Einstein-Maxwell field still remain to be explored.

Hydrodynamics	Gravitation plus electromagnetism
1. Divergence condition; pressure-density relation; and equations of motion.	1. Maxwell's source-free equations in curved space and Maxwell's stress-energy tensor as source term in Einstein's gravitational field equations.
2. Formulation in terms of a single action principle.	2. Formulation in terms of a single action principle.
3. Descriptions in alternative coordinate systems related by tensor analysis.	3. Description has same form in all coordinate systems, much as in group theory the laws of multiplication of the matrices that represent group elements are independent of the special choice of representation.
4. Hedlund and others have made a beginning at representing the laws of mechanics in an abstract coordinate-free form.[a]	4. Nothing known to have been done to represent the Maxwell-Einstein theory in a coordinate-free form as abstract as the abstract theory of groups.
5. Two alternative formulations of equations, according as one analyzes the time change of the hydrodynamic quantities at a given space point or at a given mass point	5. Coordinate system employed, x^i, may be arbitrary, or may be given an invariant significance (A. Komar) by identification with four of the fourteen invariants, I_s, of Géhéniau and Debever.[f]
6. Degree of arbitrariness in specification of initial conditions well understood.	6. Cartan-Lichnerowicz equations must be satisfied by initial conditions, but means are not yet known to generate the general solution of these initial value requirements.[g]
7. Expansions for hydrodynamic quantities near a typical point.	7. Expansion for g_{ik} and F_{ik} and R_{ijkl} near a typical point.
8. Behavior of fluid near a stagnation point or a triple point or a vortex center.	8. Behavior of electromagnetic and gravitational field near a point of special symmetry.
9. Fluid motion generated by an elemental source or sink.	9. Schwarzschild and Reissner-Nordström solution.[h]
10. Expansion of density, pressure and other quantities near a point in a series of spherical harmonics.	10. Similar expansion in spherical harmonics for small departures from a condition of spherical symmetry seems not yet to have been given.[i]
11. Sound waves; radiation pattern related to source geometry; radiation damping.	11. Electromagnetic and gravitational waves in small amplitude approximation and their radiation pattern as related to the geometry of the source. Radiation reaction analyzed for electromagnetic but not for gravitational radiation.
12. Shock waves; Mach triple point; slip stream.[b]	12. Behavior of waves of high amplitude not yet known, even qualitatively, except in very special cases.
13. Vortex ring moving in quiet fluid maintains its identity and integrity for a long time.	13. Collections of electromagnetic or gravitational waves or both under suitable conditions hold themselves together gravitationally for long periods of time ("geons").
14. Law of motion for the center of a vortex follows from the hydrodynamic equations themselves.[c]	14. Law of motion for a geon follows from the field equations themselves; the law of motion along a geodesic does not have to be introduced as a separate postulate.
15. Turbulence under appropriate conditions describable in statistical terms.	15. Radiation—isotropic or not—under appropriate conditions also describable in statistical terms.
16. Generation of sound waves by turbulence partly studied.[d]	16. Gravitation-induced interactions between electromagnetic waves, gravitational waves and geons, and cross sections for elementary types of encounter between these objects, hardly analyzed at all so far.
17. Many special solutions of the hydrodynamic equations are known from similarity arguments or group theory or other special methods of analysis.	17. Problem of homogeneous isotropic closed radiation-filled universe and its expansion and subsequent contraction, and a few other special problems have been analyzed. Field largely unexplored.[j]
18. Many types of hydrodynamic instability have been analyzed, among them Rayleigh-Taylor instability and Holmholtz instability and Bénard cells.[e]	18. One does not know how small departures from sphericity grow with time in the problem of the expanding universe, nor in the Schwarzschild-Reissner-Nordström solution. The problem of stability analysis is practically untouched.[i]
19. One has developed a set of secondary concepts adequate to describe many derived properties of the hydrodynamic field: turbulence; vorticity; acoustic impedance; radiation flux, etc.	19. Terminology for the electromagnetic field, with its six components, is extraordinarily rich; but for the gravitational field, with twenty R_{ijkl}, the present secondary conceptual structure is very rudimentary, having as yet no terms analogous to dielectric constant, permeability, Poynting vector, radiation pressure, Thomson scattering cross section, inductance, etc.

[a] G. A. Hedlund, Bull. Am. Math. Soc. **45**, 241 (1939); also G. A. Hedlund and M. Morse, Am. J. Math. **60**, 815 (1938), and W. H. Gottschalt and G. A. Hedlund, *Topological Dynamics*, Am. Math. Soc. Colloquium Publ. **36** (Providence, Rhode Island, 1955).

[b] See, for example, W. Bleakney and A. H. Taub, Revs. Modern Phys. **21**, 584 (1949); Fletcher, Bleakney, and Taub, Revs. Modern Phys. **23**, 271 (1951).

[c] See, for example, H. Lamb, *Hydrodynamics* (Cambridge University Press, London, 1953), sixth edition, Chap. VII.

[d] M. J. Lighthill, Proc. Roy. Soc. (London) **A211**, 564 (1952); *ibid.*, **A222**, 1 (1954).

[e] See, for example, reference c, Chap. XI; also Lord Rayleigh, *The Theory of Sound* (MacMillan and Company, Ltd., London, 1896), second edition, Chap. XXI; H. Bénard, Rev. gen. sci. pures appl. 12, 1261 (1900); Ann. chim. phys. **23**, 62 (1901).

[f] A. Komar, Proc. Natl. Acad. Sci. U. S. **41**, 758 (1955); and Ph.D. thesis, Princeton, New Jersey, 1956 (unpublished); T. Y. Thomas, Proc. Natl. Acad. Sci. U. S. **31**, 306 (1945); J. Géhéniau and R. Debever, Acad. roy. de Belgique **42**, 114 (1956).

[g] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson et Cie, Paris, 1955), Chap. II.

[h] H. Reissner, Ann. Phys. **50**, 106 (1916); G. Nordström, Proc. Amsterdam Acad. **21**, 68 (1918).

[i] *Note added in proof.*—The Schwarzschild solution has been shown stable against small perturbations in a paper submitted for publication by T. Regge and J. A. Wheeler. The stability of the expanding universe against small departures from sphericity has been partially treated by E. Lifshitz.

[j] See for example, L. Landau and E. Lifschitz, *The Classical Theory of Fields* (Addison-Wesley Press, Cambridge 42, Massachusetts, 1951), Chap. XI.

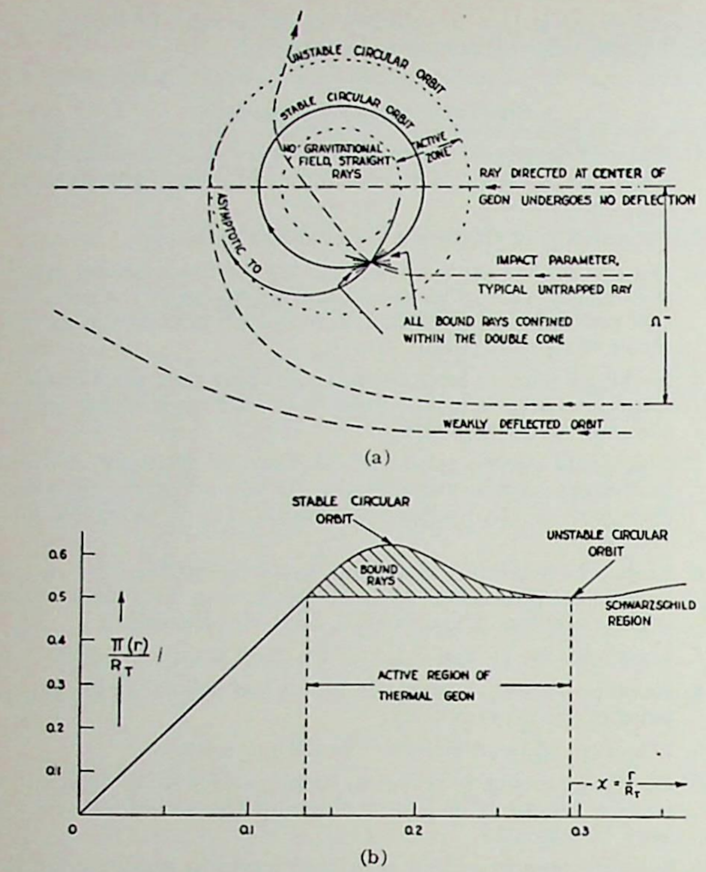


FIG. 1. (a) Schematic diagram of null geodesics or ray orbits in the gravitational field of a thermal geon. Smooth curves: bound orbits. Dashed curves: orbits that come from or go to infinity. Dotted curves: inner and outer boundaries of the active region of the geon; i.e., limits between which all bound rays circulate. (b) A ray which is to reach to the point r must have an impact parameter less than $\pi(r)$. Here $\pi(r)$ $(=re^{-\nu/2})$ is plotted as a function of r in dimensionless units. Rays that are bound must have impact parameters between $P_1=0.51R_T$ and $P_2=0.62R_T$, where R_T is a unit of length uniquely determined by the temperature.

demands localization of each mode of electromagnetic oscillation in a region where the dimensionless measure, $-g_{44}$, of gravitational potential $(=1-2\ GM/c^2r$ for Schwarzschild's point mass solution) is small compared with unity, and orientation of the flow of field energy in each mode normal to the gradient of $-g_{44}$, or in a direction not too far from normal. In contrast, a photon that travels outward along a radius vector in a spherical geon can always escape.

That region of space where $-g_{44}$ is small compared to unity, or where the rate of ticking of a standard clock is greatly slowed, is in some ways analogous to the closed container of the theory of blackbody radiation. There every mode is confined, while here only the bounded modes—or modes of semitangential energy transport—properly belong to the interior of the container. Except for this difference the varieties of geon are as numerous as the states of electromagnetic field disturbance in a hohlraum. In the general case many modes of oscillation of the field are excited, each with an amplitude of its own. Then the gravitational field resulting from the average energy distribution has no particular symmetry, and the mathematical analysis is complex. There are at least three particularly simple types of geons. (i) Toroidal geons. Here the electromagnetic energy is concentrated in the equatorial ring where the gravitational potential has its extreme value. The energy flows around the ring in equal or unequal measure in the positive and negative senses. The ratio of the fluxes in the two directions is unity or zero in the two special subcases of zero angular momentum and maximal angular momentum. Toroidal geons presumably have the greatest stability of all geons of a given mass and angular momentum. They have received partial mathematical analysis by F. J. Ernst.[2] (ii) Simple spherical geons. Here the gravitational field has nearly spherical symmetry. The electromagnetic energy is localized in a single mode of field oscillation with no radial nodes and has a large but definite number, l, of nodes in the angular variation of the field quantities. More accurately, the energy is divided over a nearly degenerate system of modes, corresponding to different spherical harmonics of the same l value but different m values. Such geons[1] require for specification two parameters: the mass and the azimuthal index number, l. The electromagnetic fields are intense in a relatively thin spherical shell, or active region. In the course of time nonlinear couplings build up the strength of originally unexcited modes, and the energy distribution slowly becomes more diffuse. (iii) Thermal geons. Here the gravitational field is again spherically symmetric, but the available electromagnetic energy is distributed over all confined modes according to the most natural of statistical laws. A thermal geon is in many ways the simplest to discuss because all its properties are fixed by a single parameter, the temperature.

Orders of magnitude and scaling laws for thermal geons are readily discussed. Let the energy density of blackbody radiation be denoted by $bT^4/\hbar^3c^3$, where $b=\pi^2/15$. The energy density in the active region is of the same order

$$\text{energy/volume}\sim Mc^2/R^3\sim bT^4/\hbar^3c^3. \tag{3}$$

The linear extension, R, of the active region is set by the requirement that $-g_{44}$ deviate substantially from unity inside:

$$GM/c^2R\sim 1. \tag{4}$$

From (3) and (4) the size is of the order

$$R\sim R_T\equiv(\hbar^3c^7/8\pi bGT^4)^{\frac{1}{2}}, \tag{5}$$

where the factor 8π has been included for later convenience; and the mass is of order

$$M\sim M_T\equiv(\hbar^3c^{11}/8\pi bG^3T^4)^{\frac{1}{2}}. \tag{6}$$

Higher temperature corresponds to geons of greater energy concentration, and of smaller mass and size.

There is a maximum temperature, and a smallest size, for which geons are free from electron pair creation and annihilation phenomena. (1) Thermal energies

[2] F. J. Ernst, Phys. Rev. 105, 1662 and 1665 (1957).

must be insufficient to creat pairs, or the temperature T must be significantly less than $2mc^2$. (2) The electric field $\mathcal{E}$ acting on an electronic charge over the characteristic localizability distance for an electron, $\hbar/mc$, must be insufficient to raise the particle to a state of positive energy:

$$\mathcal{E} < \mathcal{E}_{\text{crit}} \equiv m^2c^3/e\hbar = 4.41\times 10^{13}\ (\text{g/cm sec}^2)^{\frac{1}{2}}$$
$$(=4.41\times 10^{13}\ \text{gauss}). \quad (7)$$

Consequently the energy density of the field must be limited to values less than $\mathcal{E}_{\text{crit}}^2/8\pi$ to permit a simple analysis:

$$(\pi^2/15)T^4/\hbar^3c^3 < (m^2c^3/e\hbar)^2/8\pi$$

whence

$$T < [(15/\pi^2)(137/8\pi)]^{\frac{1}{4}}mc^2 = 1.70mc^2(1.01\times 10^{10}\ {}^\circ\text{K}).$$

Corresponding to the limit $T \sim mc^2$ on the temperature there are limits

$$\begin{aligned} &\sim (mc^2)^4/\hbar^3c^3 = 1.41\times 10^{25}\ \text{erg/cm}^3 \text{ on energy density} \\ &\sim m^4c^3/\hbar^3 = 1.57\times 10^4\ \text{g/cm}^3 \text{ on mass density} \\ &\sim (\hbar^3/Gm^4c)^{\frac{1}{2}} = 9.25\times 10^{11}\ \text{cm on size} \\ &\sim (\hbar^3c^3/G^3m^4)^{\frac{1}{2}} = 1.25\times 10^{40}\ \text{g on mass.} \end{aligned} \quad (8)$$

All analysis in this article confined to geons on the high mass side of these limits.

In the realm of sizes that are free of electron physics, geons satisfy the simple scaling law implied by (5) and (6): when one geon is hotter than another by a factor or two, it has a mass and radius four times as small; but apart from this difference of scale it has the same law of fall off of activity as does the cooler geon. For this reason the calculations are carried out in terms of the dimensionless scale independent variable

$$x = r/R_T, \quad (9)$$

where R_T is the characteristic distance of (5). It is found that the mass has the value

$$M = 0.099\ M_T, \quad (10)$$

where M_T is the characteristic mass value of (6). The total mass $M(r)$ out to the distance r, and the dimensionless measure of this mass,

$$m(x) = M(r)/M_T,$$

are *defined* by

$$e^{-\lambda} = 1 - 2GM(r)/c^2r = 1 - 2m(x)/x \quad (11)$$

The dimensionless measure of mass, $m(x)$, is shown in Fig. 2, along with the metric quantities $-g_{TT}(x) = e^\nu$ and $g_{rr} = e^\lambda$. Figures 1 and 2 contain the principal results of this paper.

The analysis is simple in outline. If the gravitational field is temporarily assumed known, then the information is at hand to set up Maxwell's equations for characteristic vibrations of the electromagnetic field. Each solution has a well-determined proper frequency Ωc, and is therefore assigned by Planck's formula a definite amount of energy. Distribution of this energy in space follows from the form of the field eigenfunction. Summation over all characteristic modes then gives the total energy density—and similarly the total stress—in the field at any specified point in space. The stress-energy density so found—or its time average value—constitutes the entire source for the original static gravitational field. One arrives at a coupled system of equations for the gravitational field and for the characteristic vibrations of the electromagnetic field.

In practice the number of characteristic modes of appreciable excitation is so enormous that a statistical treatment of the characteristic modes is more appropriate than any detailed solution of Maxwell's equations. The Fermi-Thomas picture of electrons in an atomic field bears a relation to the Schrödinger equation like the connection between the statistical analysis of optical rays and the solutions of Maxwell's equations in a gravitational field. In both cases the JWKB approximation method links the wave point of view to the particle or ray picture. In Sec. 2 Maxwell's equations in a spherically symmetric metric are separated into an angular and a radial part. The solution of the radial part in the JWKB approximation is found to depend upon a single constant, the impact parameter P. For a ray which can escape to infinity this distance is defined as the asymptotic separation of the ray and a parallel ray which comes straight through the center of the geon without deflection. For a trapped ray, the concept of ray provides an idealized description of a mode of vibration of the electromagnetic field. The amplitude is large and oscillatory in the region of trapping, falls off exponentially outside the zone of confinement, but with still greater distance starts again to oscillate, corresponding to a weak leakage wave that runs off to infinity. This leakage wave, described in ray language, has an impact parameter, P, that is a property of the mode of vibration as a whole. Bound rays in a thermal geon have impact parameters between the two limits (Fig. 1),

$$P_1 = 3^{\frac{3}{2}}GM/c^2 = 0.51R_T$$

and

$$P_2 = 0.62R_T, \quad (12)$$

and move always between the limits

$$R_{\min} = 0.14R_T \text{ and } R_{\max} = 0.30R_T. \quad (13)$$

Every point in space is characterized by a critical impact parameter,

$$\pi(r) = r\exp[-\tfrac{1}{2}\nu(r)]. \quad (14)$$

No ray with impact parameter greater than $\pi(r)$ can arrive at r. Trapped rays are associated with the shaded region in Fig. 1. Among these, the bound null geodesics, there is exactly one which is circular. Its radius is

$$R_{\text{circle}} = 0.19R_T \quad (15)$$

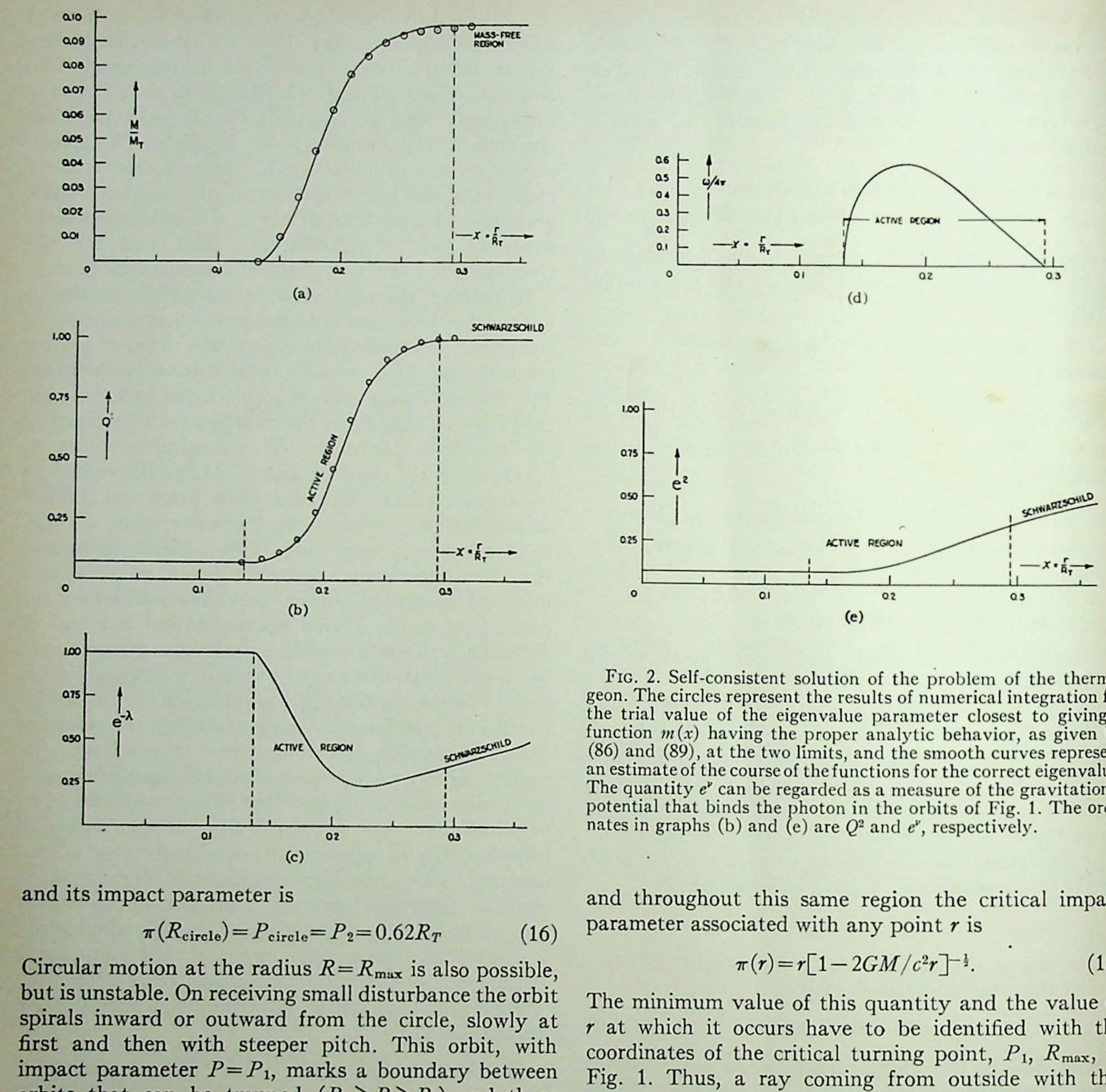


FIG. 2. Self-consistent solution of the problem of the thermal geon. The circles represent the results of numerical integration for the trial value of the eigenvalue parameter closest to giving a function $m(x)$ having the proper analytic behavior, as given by (86) and (89), at the two limits, and the smooth curves represent an estimate of the course of the functions for the correct eigenvalue. The quantity e^{ν} can be regarded as a measure of the gravitational potential that binds the photon in the orbits of Fig. 1. The ordinates in graphs (b) and (e) are Q^2 and e^{ν}, respectively.

and its impact parameter is

$$\pi(R_{\text{circle}})=P_{\text{circle}}=P_2=0.62R_T \tag{16}$$

Circular motion at the radius $R=R_{\max}$ is also possible, but is unstable. On receiving small disturbance the orbit spirals inward or outward from the circle, slowly at first and then with steeper pitch. This orbit, with impact parameter $P=P_1$, marks a boundary between orbits that can be trapped ($P_2 \geqslant P \geqslant P_1$) and those that can not.

The impact parameter $P=P_1$ gives the bound orbit with the greatest possible range of excursion in r, from $r=R_{\min}$ to $r=R_{\max}$. The bound orbit with the maximum possible impact parameter, $P=P_2$ has the minimum radial excursion: none at all.

It is easy to discuss the behavior of rays of impact parameter $P=P_1$ near the outer limit, $r=R_{\max}$, of the region of trapping. All the mass M of the system lies within the radius $R_{\max}$, and the total angular momentum of all the radiation mass-energy is zero. Consequently the metric from the point $r=R_{\max}$ to infinity has the Schwarzschild value,

$$e^{\nu}=e^{-\lambda}=1-(2GM/c^2r); \tag{17}$$

and throughout this same region the critical impact parameter associated with any point r is

$$\pi(r)=r[1-2GM/c^2r]^{-\frac{1}{2}}. \tag{18}$$

The minimum value of this quantity and the value of r at which it occurs have to be identified with the coordinates of the critical turning point, P_1, $R_{\max}$, in Fig. 1. Thus, a ray coming from outside with this critical impact parameter, like a ray coming from inside, must approach the point $r=R_{\max}$ without ever reaching it, in the manner of a spiral asymptotic to a circle. By differentiation of (18) we find

$$R_{\max}=3GM/c^2$$
$$P_1=\pi(R_{\max})=3^{\frac{3}{2}}GM/c^2. \tag{19}$$

The comparison between the two extreme bound orbits is striking (Fig. 1). One is exactly a circle of radius $r=R_{\text{circ}}$. The other, starting as a straight line tangent to the smaller circle $R_{\min}$, moves out to larger r, gradually curving as it goes, crosses the circle $r=R_{\text{circ}}$, and ends up at the circle $r=R_{\max}$ after an infinite number of convolutions.

Let one describe a sphere of a specific radius r between

$R_{\min}$ and $R_{\max}$ and ask which of the bound orbits cut this sphere. Not all of them, unless r happens accidentally to equal R_{circle}. A typical trapped ray ($P_2 \geqslant P \geqslant P_1$) will or will not reach the distance r according as its impact parameter is less than or greater than the critical impact parameter, $\pi(r)$, associated with the point r. The extremal ray $P=\pi(r)$, that can barely touch the sphere moves at *that* part of its orbit perpendicular to the axis. A ray of a slightly smaller impact parameter cuts across the sphere of radius r both on its outward excursion and on its return, both times with only a small velocity component normal to the surface. A bound ray of minimum impact parameter P_1, cuts this sphere closer to normality than does any other bound ray. Rays of this impact parameter observed at the point r will have all azimuths. Including both outgoing and reflected portions of orbits, these rays define a double cone. This double cone bounds the bundle of directions filled out by all the trapped rays ($\pi(r) \geqslant P \geqslant P_1$). This bundle embraces the largest solid angle when $r=R_{\text{circle}}$; i.e., when $\pi(r)=\pi_{\text{max bound}} = P_2$. The bundle narrows down to a flat disk perpendicular to the radius vector when r approaches either of the limiting values $R_{\min}$ or $R_{\max}$; i.e. $\pi(r)=P_1$.

For a quantitative measure of ray direction we define an angle of inclination, $\alpha=0$, for a ray (unbound!) that travels parallel to the radius vector. For any other ray the inclination, α, depends upon the impact parameter P that characterizes the whole course of that ray and the point r at which that ray is observed:

$$\tan\alpha = \frac{d\text{ (proper distance }\perp r)}{d\text{ (proper distance }\| r)}$$

$$= rd\theta/e^{\lambda/2}dr = [\pi^2(r)/P^2-1]^{-\frac{1}{2}}; \qquad (20)$$

$$\sin\alpha = P/\pi(r).$$

The bundle of bound rays has, at the point r, angles of inclination that range from

$$\alpha = \text{arc sin}[P_1/\pi(r)] \qquad (21)$$

for outgoing rays of the minimum impact parameter P_1, through $\alpha=\pi/2$ for rays of the local extremal impact parameter $\pi(r)$, to

$$\alpha = \pi - \text{arc sin}[P_1/\pi(r)] \qquad (22)$$

for returning rays of the minimum impact parameter, P_1. The allowed rays fill a fraction of the entire solid angle given by

$$\omega/4\pi = \cos\alpha = [1-P_1^2/\pi^2(r)]^{\frac{1}{2}}. \qquad (23)$$

Normal blackbody radiation would have the energy density

$$bT^4/\hbar^3c^3 = (\pi^2/15)T^4/\hbar^3c^3. \qquad (24)$$

The actual energy density is less because only a portion of the solid angle is filled with blackbody radiation (23) and because the effective temperature at the point r is not T, but

$$T/(-g_{44})^{\frac{1}{2}} = Te^{-\frac{1}{2}\nu(r)}, \qquad (25)$$

according to Tolman.[3] Thus the energy per unit volume is

$$T_T{}^T = (bT^4/\hbar^3c^3)(\omega/4\pi)e^{-2\nu}. \qquad (26)$$

The same result is obtained from first principles in §3 by evaluating the energy density of each individual bound mode of vibration of the electromagnetic field and summing over all bound modes. The JWKB approximation is used, together with the wave-ray correspondence. All summations are replaced by integrations in view of the enormous number of proper vibrations of appreciable energy content. A similar calculation gives for the radial component of the electromagnetic stress energy tensor

$$T_r{}^r = (1/3)(bT^4/\hbar^3c^3)(\omega/4\pi)^3e^{-2\nu}. \qquad (27)$$

The circumstance that one has in principle to deal with a practically infinite number of eigenvalue problems in treating the trapped modes of vibration, far from causing difficulties, helps to express the stress-energy tensor in terms of purely geometrical quantities.

In Sec. 4 the self-consistent system of equations of the geon is analyzed and solved. Thus, (26) and (27) with (23) and (14) give the source of the gravitational field—the electromagnetic stress energy tensor—in terms of the gravitational field itself. In addition we require only the law to find the gravitational field from its sources—Einstein's generalization of the equation of Newtonian theory,

$$\nabla^2\varphi = 4\pi G\rho, \qquad (28)$$

where φ is Newtonian potential and ρ is the mass density.

The field equations of general relativity give the second derivatives of the g_{ik} in terms of the ten source strengths T_{ik}, of which the component, T_{44}, divided by c^2 is the mass density. The metric tensor, g_{ik}, is completely known except for the two dilatation functions, $\lambda(r)$ and $\nu(r)$. Consequently, two of the ten equations are sufficient to determine the potentials in terms of the source strengths:

$$R_T{}^2[e^{-\lambda}(r^{-1}d\nu/dr+r^{-2})-r^{-2}] = (8\pi GR_T{}^2/c^4)T_r{}^r = (\tfrac{1}{3})(\omega/4\pi)^3e^{-2\nu}, \qquad (29)$$

$$R_T{}^2[e^{-\lambda}(r^{-2}-r^{-1}d\lambda/dr)-r^{-2}] = (8\pi GR_T{}^2/c^4)T_T{}^T = -(\omega/4\pi)e^{-2\nu}. \qquad (30)$$

Here the expressions (26) and (27) for the energy stress tensor in terms of the solid angle have been used. In terms of the dimensionless independent variable,

[3] R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Clarendon Press, Oxford, 1934), p. 318. T corresponds to Einstein's "wahre Temperatur"; $Te^{-\frac{1}{2}\nu}$ to Ehrenfest's "Taschen-temperatur."

$x=r/R_T$; these are

$$e^{-\lambda}(x^{-1}d\nu/dx+x^{-2})-x^{-2}=(\tfrac{1}{3})(\omega/4\pi)^3e^{-2\nu} \\ =(\tfrac{1}{3})(1-x_1^2x^{-2}e^{\nu})^{\frac{3}{2}}e^{-2\nu} \quad (31)$$

$$e^{-\lambda}(x^{-2}-x^{-1}d\lambda/dx)-x^{-2}=-(\omega/4\pi)e^{-2\nu} \\ =-(1-x_1^2x^{-2}e^{\nu})^{\frac{1}{2}}e^{-2\nu}. \quad (32)$$

These are the two differential equations for the self-consistent solution of the thermal geon problem.

In (31) and (32) the quantity x_1 is an eigenvalue parameter which measures the critical impact parameter P_1 in dimensionless units. Also it determines what mass M is required to give stability to a thermal geon of temperature T:

$$x_1\equiv P_1/R_T=(3^{\frac{3}{2}}GM/c^2)/(GM_T/c^2)=3^{\frac{3}{2}}M/M_T. \quad (33)$$

This parameter has to be chosen in such a way as to yield a solution that is acceptable in the following sense:

(1) The metric dilatation parameters must join on smoothly to the Schwarzschild values at the outer boundary of the active region; (2) The metric must become flat at the inner boundary of the active region. Specifically, from Eqs. (17), (18), and (19), the conditions at the outer join point, $x=x_{\max}$, are

$$x_{\max}=x_1/3^{\frac{1}{2}}, \\ e^{\nu}=\tfrac{1}{3};\quad d\nu/dx=3^{\frac{1}{2}}(2/x_1), \quad (34) \\ e^{\lambda}=3;\quad d\lambda/dx=-3^{\frac{1}{2}}(2/x_1).$$

At the inner join point, $x=x_{\min}$, the boundary conditions are

$$(xe^{-\frac{1}{2}\nu})\text{ at }x_{\min}=x_1\text{ (definition of }x_{\min}) \\ d\nu/dx=d\lambda/dx=0. \quad (35)$$

For convenience in numerical integration of the self-consistent equations these boundary conditions on $\lambda(x)$ were expressed as boundary conditions on a variable, $m(x)$—the effective mass up to a distance x, defined by the question

$$e^{-\lambda}=1-2m(x)/x. \quad (36)$$

At $x_{\max}=x_1/3^{\frac{1}{2}}$ we have $m(x_{\max})=(\frac{1}{3})x_{\max}$ and $(dm/dx)=0$; at the inner point both $m(x)$ and (dm/dx) are zero. The differential equations were integrated with an electronic digital computer starting at a specific inner join point and proceding outwards until either $m(x)=x/3$ or $dm/dx=0$. In the former case the deviation from flatness was noted, the inner join point made larger, and the integration repeated from the start. When $m(x)$ became flat before reaching the value $(\frac{1}{3})x_{\max}$, the inner join point was made smaller and again the integration repeated. Figures 1 and 2 report the results of the integration for the best value of x_1 found with the limited machine time available

$$x_1=0.51. \quad (37)$$

Details of the present analysis of thermal geons follow.

2. THE PROPER VIBRATIONS OF THE ELECTROMAGNETIC FIELD

Analysis In Spherical Harmonics

The electromagnetic field $F_{jk}(F_{23}\sim H_x;\ F_{14}\sim E_x)$ satisfies the eight source-free Maxwell equations

$$(-g)^{-\frac{1}{2}}(\partial/\partial x^{\alpha})(-g)^{\frac{1}{2}}F^{i\alpha}=0 \quad (38)$$

and

$$\{i\alpha\beta\gamma\}\partial F_{\alpha\beta}/\partial x^{\alpha}=0. \quad (39)$$

Here the numbers $\{ijkl\}$ are defined by $\{1234\}=1$ and by $\{ijkl\}$ changing sign on reversal of any two indexes. Also g is the determinant of the g_{ik}: $g=-r^4\sin^2\theta \times\exp[\lambda(r)+\nu(r)]$. In the static spherically symmetric gravitation field the Maxwell equations are invariant with respect to a group of transformations built from the following elementary operations: (1) rotation of the space frame of reference; (2) translation of time; (3) inversion of the space frame in the center of symmetry; (4) reflection in a plane through this center; (5) reversal of time; (6) interchange of the roles of electric and magnetic fields, according to the substitution

$$F_{jk}{}^{II}=\tfrac{1}{2}(-g)^{\frac{1}{2}}\{ik\alpha\beta\}g^{\alpha\mu}g^{\beta\nu}F_{\mu\nu}{}^{I}, \quad (40)$$

such that $E^{II}\sim H^{I}$, $H^{II}\sim -E^{I}$. For a flat space time continuum one knows the irreducible representations of this group of transformations and the basic set of functions on which these transformations operate,[4] and it is very easy to modify the results for the present spherically symmetrical curved static space time continuum. The basic solutions are characterized by a circular frequency, Ω/c, by two angular momentum quantum numbers, l and m, and by the statement that the disturbance is of electric or magnetic multipole character, as the case may be, in the following sense: (1) Magnetic multipole field: the electric field is everywhere exactly perpendicular to the radius vector; the parity of the electric field with respect to space inversion is $(-1)^l$; the parity of the magnetic field is $(-1)^{l+1}$; and the field is expressed in terms of a four potential A_i

$$F_{jk}=\partial A_k/\partial x^j-\partial A_j/\partial x^k, \quad (41)$$

which lies always on the surface of a sphere with center at the origin, and points always in the direction of the electric field. Thus $A_r=A_T=0$, and the space vector $\mathbf{A}$ is a function of r multiplied by $\mathbf{r}\times\nabla$ acting on a spherical harmonic. The normalized spherical harmonic[4] is denoted by $Y_l^{(m)}(\theta,\phi)$. We write

$$Q\equiv Y_l^{(m)}(\theta,\phi)\exp(-i\Omega T); \quad (42)$$

then the nonzero components of the vector potential

[4] See, for example, J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley and Sons, Inc., New York, 1952), pp. 799 ff. for formulas and references to the original literature.

for the mode in question have the form

$$A_\theta = a(r) imQ/\sin\theta + \text{complex conjugate},$$
$$A_\phi = -a(r)\sin\theta\partial Q/\partial\theta + \text{complex conjugate}, \quad (43)$$

where the radial function $a(r)$ has the dimensions $(\text{g cm}^3/\text{sec}^2)^{\frac{1}{2}}$ or gauss cm^2 or cm electrostatic volts. The field itself, F_{jk}, is expressed relative to the coordinate system that appears in the metric (1). However, the field is more easily visualized in a local orthonormal coordinate system that is aligned along the r, θ, ϕ, T axes. In this frame we shall call the field (**E**,**H**). From (43) we derive

$$\begin{aligned}
e^{\nu/2+\lambda/2}E_r &= F_{rT} = 0\\
re^{\nu/2}E_\theta &= F_{\theta T} = -\Omega amQ/\sin\theta + \text{c.c.}\\
r\sin\theta e^{\nu/2}E_\phi &= F_{\phi T} = -\Omega ai\sin\theta\cdot\partial Q/\partial\theta + \text{c.c.}\\
r^2\sin\theta H_r &= F_{\theta\phi} = l(l+1)a\sin\theta Q + \text{c.c.}\\
r\sin\theta e^{\lambda/2}H_\theta &= F_{\phi r} = (da/dr)\sin\theta\cdot\partial Q/\partial\theta + \text{c.c.}\\
re^{\lambda/2}H_\phi &= F_{r\theta} = (da/dr)imQ/\sin\theta + \text{c.c.}
\end{aligned} \quad (44)$$

(2) Electric multipole field: the magnetic field is everywhere exactly perpendicular to the radius vector; space inversion multiplies the electric field by $(-1)^{l+1}$, the magnetic field by $(-1)^l$. This type of field is obtained from the magnetic multipole field by two steps: (a) replace $a(r)$ in (43) by $b(r)$ and call the field F so calculated "F^{I}" (b) substitute this field into (40) and calculate the field "F^{II}." This is the desired electric multipole field:

$$\begin{aligned}
F_{rT} &= l(l+1)\Omega^2 r^{-2} e^{\lambda/2+\nu/2} b(r)Q + \text{c.c.}\\
F_{\theta T} &= \Omega e^{-\lambda/2+\nu/2}(db/dr)(\partial Q/\partial\theta) + \text{c.c.}\\
F_{\phi T} &= \Omega e^{-\lambda/2+\nu/2}(db/dr)imQ + \text{c.c.}\\
F_{\theta\phi} &= 0\\
F_{\phi r} &= e^{\lambda/2-\nu/2}b(r)mQ + \text{c.c.}\\
F_{\theta r} &= e^{\lambda/2-\nu/2}b(r)i(\partial Q/\partial\theta) + \text{c.c.}
\end{aligned} \quad (45)$$

Radial Function

The radial factors $a(r)$ and $b(r)$ in the field expressions (44) and (45) satisfy a common differential equation,[5]

$$d^2f/dr^{*2} + [\Omega^2 - l(l+1)e^\nu/r^2]f = 0 \quad (46)$$

where

$$dr^* = e^{\lambda/2-\nu/2}dr.$$

We demand a solution of (46) that is regular at the origin. Near the origin λ and ν are constant, and λ is zero.[1] Consequently, f varies as r^{l+1} near the origin. At larger r, f increases roughly exponentially until the first zero of the square bracket in (46): $r=r_1$. Beyond this point the quantity

$$l(l+1)e^\nu/\Omega^2r^2 \quad (47)$$

falls below unity, the square bracket is positive, and the solution is oscillatory. With further increase in r there are two possibilities which are familiar from the closely analogous problem of alpha decay: (1) The quantity (47) never rises above unity again. In this case the solution remains oscillatory to infinite r, ultimately approaching the behavior

$$f(r) \sim c_1\sin(\Omega r + c_2). \quad (48)$$

Such a solution represents an electromagnetic wave that runs freely to infinity (analogous to an alpha particle with energy that exceeds the potential barrier). Since such a wave carries energy away from the geon, it is not of interest in constructing a relatively stable object. Such modes of oscillation of the electromagnetic field are assigned zero energy. (2) The quantity (47) rises above unity again at a point $r=r_2$ (analogous to inner radius of potential barrier for an alpha particle with energy below the top of the barrier). A sufficiently great increase in r will result in (47) once again falling below unity at some point, $r=r_3$. The solution, $f(r)$, of (46) rises monotonically from the origin to r_1, oscillates from r_1 to r_2, behaves between r_2 and r_3 as a linear combination of a function that falls roughly exponentially and another function that rises roughly exponentially, and resumes an oscillatory character from r_3 to ∞. The wave in this outer region transports energy away from the geon. The rate of transport is the smaller, the less is the ratio of the amplitudes of oscillation in the outer and inner regions. This ratio has a minimum value for certain characteristic values of Ω, designated as Ω_{nl}. At such a value of Ω the solution $f(r)$ decreases monotonically and roughly exponentially all the way from r_2 to r_3. For such characteristic solutions the region from $r_1(\Omega_{nl})$ to $r_2(\Omega_{nl})$ is called the region of activity. As a first approximation exponential fall-off is supposed to continue indefinitely beyond the point r_2.

Eigenvalues, Eigenfunctions, and Averages

The characteristic values, Ω_{nl}, of Eq. (46) in the sense just defined, are given in the JWKB approximation by the implicit equation,

$$\int_{r_1}^{r_2}[1 - l(l+1)e^\nu/\Omega_{nl}^2r^2]^{\frac{1}{2}}\Omega_{nl}dr^* = (n+\tfrac{1}{2})\pi. \quad (49)$$

The solutions themselves[6] in the same approximation have the form

$$f_{nl}(r) = C_{nl}[1 - l(l+1)e^\nu/\Omega_{nl}^2r^2]^{-\frac{1}{4}}$$
$$\cdot\sin\left\{\int_{r_1}^{r}[1 - l(l+1)e^\nu/\Omega_{nl}^2r^2]^{\frac{1}{2}}\Omega_{nl}dr^* + \frac{\pi}{4}\right\}. \quad (50)$$

[5] Reference 1, p. 520. See also A. Eddington, *The Mathematical Thoery of Relativity* (Cambridge University Press, 1923), p. 175, where the vector wave equation is written in the form $\Box A^m + R_\alpha{}^m A^\alpha = 0$, with $\Box\psi \equiv g^{\alpha\beta}\psi_{;\alpha\beta}$ (equivalent to our $(\partial/\partial x^\alpha)\times(-g)^{\frac{1}{2}}F^{m\alpha} = 0$).

[6] See, for example W. Pauli, *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1933), second edition, Vol. 24, Part 2, p. 171.

The value of $|f|^2$, averaged over a region appreciable in comparison with one wavelength, but very small relative to the size of the region of activity, is

$$\langle|f|^2\rangle \doteq \tfrac{1}{2}|C_{nl}|^2[1-l(l+1)e^{\nu}/\Omega_{nl}^2r^2]^{-\frac{1}{2}} \quad (51)$$

inside the active region, and practically zero outside. This measure of intensity, like the limits of activity, depends, not upon l and Ω individually, but only upon these two quantities in the single combination of an *impact parameter*,

$$P\equiv[l(l+1)/\Omega^2]^{\frac{1}{2}}. \quad (52)$$

At a given point in the active region of the geon this quantity can have any value between the minimum value P_1 for binding (Fig. 1) and the maximum value, $\pi(r)$ $(=re^{-\nu/2})$, appropriate to that point.

The number of bound modes of vibration that contribute to the electromagnetic field intensity at the point r in the active region is obtained by taking the product of the following factors and summing:

2,	for polarizations
$2l+1$,	for values of m
dl,	for values of l
dn,	for values of n.

Summations over l and n are replaced by integrations and then by integrations over all the relevant values of the impact parameter P between P_1 and $\pi(r)$ and over the circular cofrequency Ω by virtue of equations (30), and (33). One finds that

$$dldn=\frac{\partial(l,n)}{\partial(P,\Omega)}=\pi^{-1}\Omega\int[1-P^2e^{\nu}/r^2]^{-\frac{1}{2}}dr^*dPd\Omega,$$

and consequently the total number of modes within specified limits of Ω and P is

$$dN=(2/\pi)\Omega^2d\Omega d(P^2)\int[1-P^2e^{\nu}/r^2]^{-\frac{1}{2}}dr^*. \quad (53)$$

It is of interest to have not only the local average (51) of $|f|^2$, but also the local average of the square of its derivative

$$\langle|df/dr^*|^2\rangle \doteq \tfrac{1}{2}|C_{nl}|^2\Omega_{nl}^2[1-l(l+1)e^{\nu}/\Omega_{nl}^2r^2]^{\frac{1}{2}}. \quad (54)$$

The symbol $\langle\langle\ \rangle\rangle$ denotes an average with respect to time and with respect to position over a spherical shell of thickness large compared to a typical wave length ($\sim10^{-11}$ cm) but small compared to the dimension of the active region ($>\sim10^{11}$ cm). We note

$$\langle\langle|Y_l^{(m)}|^2\rangle\rangle=\frac{1}{4\pi};$$

$$\langle\langle|\partial Y_l^{(m)}/\partial\theta|^2+m^2\sin^{-2}\theta|Y_l^{(m)}|^2\rangle\rangle=l(l+1)/4\pi. \quad (55)$$

For a typical mode of the form (43) and (44),

$$\left.\begin{array}{r}\langle\langle F_{\theta\phi}F^{\theta\phi}\rangle\rangle=P^2e^{\nu}/r^2\\ \langle\langle F_{\phi r}F^{\phi r}+F_{r\theta}F^{r\theta}\rangle\rangle=(1-P^2e^{\nu}/r^2)\\ \langle\langle F_{\theta T}F^{\theta T}+F_{\phi T}F^{\phi T}\rangle\rangle=-1\end{array}\right\}\begin{array}{l}\text{times}\\ |C_{nl}|^2\{l(l+1)/4\pi r^2\}\\ \Omega^2e^{-\nu}[1-P^2e^{\nu}/r^2]^{-\frac{1}{2}}.\end{array}$$

The sum, $\frac{1}{2}F_{\alpha\beta}F^{\alpha\beta}=H^2-E^2$, of the quantities on the left represents the action invariant for the field oscillator. The action disappears on the time average for this oscillator as it does for every harmonic oscillator.[7]

3. ENERGY-DENSITY STRESS-DENSITY TENSOR

Radial Stress and Energy Density of One Mode

The source of the gravitational field is the stress-energy tensor of the electromagnetic field, with components

$$T_i^{\ k}=(4\pi)^{-1}F_{i\alpha}F^{k\alpha}-(16\pi)^{-1}\delta_i^{\ k}F_{\alpha\beta}F^{\alpha\beta} \quad (56)$$

so that in flat space time $-T_{44}=T_4^{\ 4}=-(E^2+H^2)/8\pi$. In a geon with a negligible rate of leakage of energy we can assume time symmetry as well as spherical symmetry. In the r, θ, ϕ, T system of coordinates, all off-diagonal elements of the stress energy tensor vanish. Among the diagonal elements there exists the equality $T_\theta^{\ \theta}=T_\phi^{\ \phi}$, and also a general relation expressing these two tangential tensions in terms of the radial tensions—in terms of $T_r^{\ r}$, $T_T^{\ T}$, and $dT_r^{\ r}/dr$. This knowledge of $T_\theta^{\ \theta}$ and $T_\phi^{\ \phi}$ does not help in the determination of the gravitational field from the field equations

$$G_i^{\ k}=(8\pi G/c^4)T_i^{\ k} \quad (57)$$

because there is an identity between $G_\theta^{\ \theta}$, $G_\phi^{\ \phi}$ and $G_r^{\ r}$, $G_T^{\ T}$, and $dG_r^{\ r}/dr$, which makes them automatically satisfy the same radial-tangential equilibrium condition as the corresponding components of the T's. For this reason it is enough to consider only the two remaining components of the stress-energy tensor, $T_r^{\ r}$ and $T_T^{\ T}$.

The intensity of the particular mode of field oscillation characterized in (55) results in this mode contributing to the energy density and radial stress:

$$\begin{aligned}&\langle\langle T_T^{\ T}\text{ (one mode)}\rangle\rangle\\&=(8\pi)^{-1}\langle\langle F_{rT}F^{rT}+F_{\theta T}F^{\theta T}+F_{\phi T}F^{\phi T}\\&\qquad-F_{\theta\phi}F^{\theta\phi}-F_{\phi r}F^{\phi r}-F_{r\theta}F^{r\theta}\rangle\rangle\\&=-(\Omega^2/16\pi^2r^2)e^{-\nu}l(l+1)|C_{nl}|^2[1-P^2e^{\nu}/r^2]^{-\frac{1}{2}}\\&\langle\langle T_r^{\ r}\text{ (one mode)}\rangle\rangle\\&=(8\pi)^{-1}\langle\langle F_{\phi r}F^{\phi r}+F_{r\theta}F^{r\theta}+F_{rT}F^{rT}\\&\qquad-F_{\theta\phi}F^{\theta\phi}-F_{\theta T}F^{\theta T}-F_{\phi T}F^{\phi T}\rangle\rangle\\&=(\Omega^2/16\pi^2r^2)e^{-\nu}l(l+1)|C_{nl}|^2[1-P^2e^{\nu}/r^2]^{\frac{1}{2}}.\end{aligned} \quad (58)$$

Many Modes; Total Mass

The electromagnetic stress energy tensor being bilinear in the field, the fields of all individual standing

[7] The integral of E^2-H^2 over all space for a field that is a superposition of two modes gives zero for the integral of the cross term in virtue of the orthogonality that follows from the eigenvalue equation (46) for radial functions of the same l and different n.

waves must be added before commencing the evaluation. For the time average value of the energy density, we assume that the elementary disturbances are incoherent[1]: the average contribution from cross terms between modes vanishes. Thus, we write

$$\langle T_T{}^T\rangle = \sum_{n,l,m,p} \langle\langle T_T{}^T(n,l,m,p\,;r,\theta,\phi)\rangle\rangle \qquad (59)$$

where the sum goes over all values of the angular index, l, and m; over all the values of the index number n of the radial proper function f_{nl} associated with the confrequency Ωnl; and over both types of polarization: p=electric or magnetic multipole. There is a similar expression for the radial stress.

There are two simple ways to get to the total mass.[8]

(1) Write

$$e^{-\lambda} = 1 - 2GM(r)/c^2 r \qquad (60)$$

as a definition of the effective mass, $M(r)$; insert this expression into differential equation (30) to derive an expression for $dM(r)/dr$; integrate from $r=0$ to $r=\infty$; and find

$$Mc^2 = c^2 M(r=\infty) = -\int_0^\infty \sum_{\text{modes}} \times\langle\langle T_T{}^T(\text{one mode})\rangle\rangle 4\pi r^2 dr. \qquad (61)$$

(2) Write down the expression for the local density of *gravitational plus electromagnetic* energy; integrate over all space; use the fact that the gravitational part is expressible in terms of a surface integral; use also the fact that this surface integral does not involve any details of the gravitational field when the metric is asymptotically flat; finally, use the fact that the trace of the electromagnetic stress-energy tensor is zero:

$$Mc^2 = \int\int\int \langle T_r{}^r + T_\theta{}^\theta + T_\phi{}^\phi - T_T{}^T\rangle(-g)^{\frac{1}{2}} dr d\theta d\phi$$

$$= -2\int \sum_{\text{modes}} \langle\langle T_T{}^T(\text{one mode})\rangle\rangle e^{\lambda/2+\nu/2} 4\pi r^2 dr. \qquad (62)$$

Expression (62) is more convenient than (61) because the radial integral is easier to calculate.

Contraction Effect

To define the energy of one mode, one can think of measuring the gravitational field far away from a geon before and after the stilling of that particular vibration. One might be tempted to write for this change in mass

$$c^2 \Delta M = I_{\text{mode}}, \qquad (63)\ (\text{wrong})$$

where

$$I_{\text{mode}} \equiv -2\int \langle\langle T_T{}^T(\text{mode})\rangle\rangle e^{\lambda/2+\nu/2} 4\pi r^2 dr. \qquad (64)$$

This would be wrong. As the radiation in question leaks out of the geon, less pressure is available to sustain the geon against gravitational forces. Consequently the geon contracts. Thus every other mode finds itself more tightly confined. Its frequency rises. Accordingly by the principle of adiabatic invariance there is a proportional *increase* in energy. Therefore, the mass of the geon will decrease by an amount less than given by (63).

The fractional change in mass due to stilling of a single mode is fantastically small. For a first primitive analysis the effect on the geon is primarily a scale transformation to smaller size. Mass and radius transform in parallel. Frequency goes up inversely as radius, and therefore inversely as mass. The same applies—according to adiabatic invariance—to the contribution, I_{mode}, to the energy from every mode except the selected mode:

$$\frac{\Delta I_{\text{mode}}\ (\text{increase})}{I_{\text{mode}}} = \frac{\Delta M\ (\text{decrease})}{M} \qquad (65)$$

(see qualification below). On this basis we compare the geon before,

$$Mc^2 = I_{\text{one}} + \sum I_{\text{others}} \qquad (66)$$

and after

$$(M-\Delta M)c^2 = 0 + \sum\left(1+\frac{\Delta M}{M}\right) I_{\text{others}}, \qquad (67)$$

the mode is stilled. Subtraction gives

$$c^2\Delta M = \frac{I_{\text{one}}}{1+(1/Mc^2)\sum I_{\text{others}}} = \tfrac{1}{2} I_{\text{one}}. \qquad (68)$$

We are considering a classical geon that has a great many proper modes, $s=1, 2, \cdots$, where s stands for the quartet (n,l,m,p). Each mode has a reduced action variable J_s (=action/2π) associated with it. The energy consists of two parts: electromagnetic energy of individual modes and gravitational energy of interaction between modes. The sum of both energies (61) is a function of the action variables of all these oscillators:

$$c^2 M = c^2 M\ (J_1, \cdots J_s, \cdots).$$

The circular frequency of a given oscillator is given by

$$c\Omega_s(J) = c^2 \frac{\partial M(J)}{\partial J_s}. \qquad (69)$$

If a particular action variable decreases slowly from its normal value to zero, during this change all other action variables keep their original values, according to the principle of adiabatic invariance.

[8] R. C. Tolman, Phys. Rev. **35**, 875 (1930); see also L. Landau and E. Lifschitz, *The Classical Theory of Fields*, translated by M. Hammermesh (Addison-Wesley Press, Cambridge, Massachusetts, 1951), pp. 309 and 323.

Then the change in mass is

$$c^2\Delta M = c\Omega_s(J)dJ_s = \tfrac{1}{2}I_s$$

$$= -\int \langle\langle T_T{}^T \text{ (chosen mode } s)\rangle\rangle e^{\lambda/2+\nu/2}4\pi r^2 dr. \quad (70)$$

A weakness in the foregoing reasoning is that it assume that stopping of the given vibration produces a change in the field of gravitational force that is represented with sufficient accuracy simply by a change of scale. Perhaps the change in metric field will act in different ways on different modes, raising some in I value less than others. It is at least conceivable that such a differentation might take place. Since the quantity of interest is not the I value of one mode, but the sum of the changes in I values of all the modes, we need, not (65) for every mode individually, but only a weaker equation for this sum of changes:

$$\frac{\sum I_{\text{ others}}(\text{increase})}{I_{\text{others}}} = \frac{\Delta M \text{ (decrease)}}{M}. \quad (71)$$

The question about correctness of (71) could be answered by an appeal back to first principles:

(1) Determine the effect on a single mode of an adiabatic change in the metric field. (a) Insert in the action principle of the electromagnetic field a trial function of the product form (43), where however the time factor, $\exp(-i\Omega T)$, is replaced by an arbitrary undetermined function of cotime, $\psi(T)$. (b) From the variational principle derive the second-order differential equation for this function. (c) Solve by the JWKB method. (d) This gives a general expression for properties of a single mode under arbitrary adiabatic changes in the Schwarzschild metric, containing one arbitrary amplitude constant, independent of both r and T.

(2) Calculate the stress energy associated with this mode, and find all over again the expressions (57), with the difference that not C_{nl}, but D_{nl}, is constant under adiabatic changes, where

$$C_{nl} = D_{nl}\Omega_{nl}{}^{-1}\left[\int [1-P^2e^\nu/r^2]^{-\frac{1}{2}}dr^*\right]^{-1} \quad (72)$$

(3) Write down the equations (29) and (30) of the self-consistent geon field for two problems: (a), with all modes s excited to specific metric-independent amplitudes D_s; and (b) with all modes but one so excited.

(4) By differencing these two sets of nonlinear equations derive linear equations for the small changes $\delta\lambda(r)$, $\delta\nu(r)$, produced in the metric by quenching one mode.

(5) Derive expressions for these changes in the metric.

(6) Find how these changes alter the I value of each mode individually, and hence

(7) Compute the change in mass of the system as a whole due to stopping one vibration.

Test of Adiabatic Analysis against Ehrenfest-Tolman Formula

For thermal geons there is an extra simplification with which we can check the correctness of (68), without the more detailed analysis just outlined. Considering blackbody radiation in equilibrium in a gravitational field for an asymptotic temperature T, we know from Ehrenfest and Tolman[9] that the energy density has the value appropriate to the temperature $T(-g_{44})^{-\frac{1}{2}}$, rather than T itself. Here this means an energy density proportional to $T^4e^{-2\nu}$, multiplied by the fractional solid angle filled by trapped rays. This result comes from the normalization (70).

The energy leaked out of the geon by quenching one mode is equated to the thermal energy of a harmonic oscillator, as given by Planck's formula:

$$\hbar c\Omega_{nl}[-1+\exp(\hbar c\Omega_{nl}/T)]^{-1} = c^2\Delta M = \tfrac{1}{2}I_s$$

$$= -\int \langle\langle T_T{}^T \text{ (chosen mode)}\rangle\rangle e^{\lambda/2+\nu/2}4\pi r^2 dr$$

$$= (\Omega^2/4\pi)l(l+1)|C_{nl}|^2 \int [1-P^2e^\nu/r^2]^{-\frac{1}{2}}dr^*. \quad (73)$$

Solving this for the amplitude $|C_{nl}|$, and inserting this value of $|C_{nl}|$ into Eqs. (38) we find *the energy density and radial stress due to one normalized mode*:

$$\langle\langle -T_T{}^T(\text{or} + T_r{}^r)\rangle\rangle = \hbar c\Omega_{nl}[-1+\exp(\hbar c\Omega_{nl}/T)]^{-1}$$

$$\cdot[1-P^2e^\nu/r^2]^{\{-\frac{1}{2}(\text{or} +\frac{1}{2})\}}(e^{-\nu}/4\pi r^2)$$

$$\times\left\{\int [1-P^2e^\nu/r^2]^{-\frac{1}{2}}dr^*\right\}^{-1}. \quad (74)$$

We multiply (74) by the number of modes, dN, as in (53), in the interval $d\Omega dP$, and sum over all modes by integrating over the wave number Ω and the equivalent impact parameter P. This gives the local value of the total energy density and radial stress:

$$\langle\langle -T_T{}^T(\text{or} + T_r{}^r)\rangle\rangle_{\text{local total}}$$

$$= \int \hbar c\Omega[-1+\exp(\hbar c\Omega/T)]^{-1}\Omega^2 d\Omega(\tfrac{1}{2}e^{-\nu}/r^2)$$

$$\times \int_{P_1}^{P_{\max}=re^{-\nu/2}} [1-P^2e^\nu/r^2]^{\{-\frac{1}{2}(\text{or} +\frac{1}{2})\}}d(P^2)$$

$$= (\pi^2T^4/15\hbar^3c^3)e^{-2\nu}\{[1-P_1{}^2e^\nu/r^2]^{\frac{1}{2}}$$

$$(\text{or } \tfrac{1}{3}[1-P_1{}^2e^\nu/r^2]^{\frac{3}{2}})\}. \quad (75)$$

In terms of the solid angle, ω, spanned by the trapped rays, the complete stress energy tensor has the form

[9] R. C. Tolman and P. Ehrenfest, Phys. Rev. 36, 1791 (1930); see also Tolman, reference 3, p. 318.

$$\langle\langle T_j{}^i\rangle\rangle_{\text{local total}}=(\pi^2T^4e^{-2\nu}/15\pi^3c^3)\times\left|\begin{array}{cccc}\frac{1}{3}(\omega/4\pi)^3 & & & \\ & \frac{1}{2}(\omega/4\pi)-\frac{1}{6}(\omega/4\pi)^3 & & \\ & & \frac{1}{2}(\omega/4\pi)-\frac{1}{6}(\omega/4\pi)^3 & \\ & & & -(\omega/4\pi)\end{array}\right|. \quad (76)$$

The stress-energy tensor calculated here by consideration of individual modes agrees with the usual blackbody value, corrected properly as demanded by the Ehrenfest-Tolman argument and by the solid angle factor. No such agreement would have resulted if the factor $\frac{1}{2}$ in (70) for the loss of mass on stilling of one mode is left out. Thus the contraction effect is essential.

Expressions (76) as derived here were used as described in the introduction to set up (29) and (30) for the self-consistent thermal geon field.

To restate the results of the present analysis, the total mass of the geon is given by the superficially paradoxical formula,

$$Mc^2=\frac{1}{2}\sum_{\text{all modes}}\begin{pmatrix}\text{Planck formula for energy}\\ \text{of one mode, as in (2)}\end{pmatrix}. \quad (77)$$

4. NUMERICAL SOLUTION OF THE DIFFERENTIAL EQUATIONS

In the dimensionless variable $x=r/R_T$ the equations for the geon field are

$$e^{-\lambda}(x d\nu/dx+1)-1=(x^2/3)(1-x_1{}^2e^\nu/x^2)^{3/2}e^{-2\nu} \quad (78)$$

$$e^{-\lambda}(1-x d\lambda/dx)-1=-x^2(1-x_1{}^2e^\nu/x^2)^{1/2}e^{-2\nu}. \quad (79)$$

Let $m(x)$ represent the effective mass, in units M_T, out to the distance $r=xR_T$, as *defined* by

$$m(x)=x(1-e^{-\lambda})/2. \quad (80)$$

Thus the mass of the geon is

$$M=m(\infty)M_T.$$

Also define a scale factor, Q, by the equation

$$Q^2=e^{\lambda+\nu}. \quad (81)$$

These variables behave as follows (Fig. 2). From $x=0$ to the radius $x=x_{\min}$, the dimensionless mass variable, m, is zero. Then it increases up to

$$m=m(\infty)=x_{\max}/3=x_1/3^{3/2} \quad (82)$$

at the outer boundary

$$x_{\max}=x_1/3^{1/2} \quad (83)$$

of the active region. Here x_1 is a measure of the critical impact parameter for trapped rays:

$$x_1=P_1/R_T. \quad (84)$$

Thereafter it remains constant at the value $m(\infty)$. The curve for m as a function of x has a horizontal slope at $x=x_{\min}$ and $x=x_{\max}$. A similar description applies to Q, with these exceptions: Q is not zero, but has a constant value between 0 and 1 for $x<x_{\min}$; and Q has the constant Schwarzschild value of unity for $x>x_{\max}$. Between these limits these quantities satisfy the equations

$$dm/dx=x^2[1-x_1{}^2x^{-2}Q^2(1-2m/x)]^{1/2}\div 2Q^4(1-2m/x)^2 \quad (85)$$

$$dQ^2/dx=x\{[1-x_1{}^2x^{-2}Q^2(1-2m/x)]^{1/2}+[1-x_1{}^2x^{-2}Q^2(1-2m/x)]^{3/2}/3\}\div Q^2(1-2m/x)^3. \quad (86)$$

Near the outer limit, $x_{\max}=x_1/3^{1/2}$, of the zone of trapping, write

$$x=x_{\max}(1+s), \quad (87)$$

where s is understood to be a small *negative* quantity. Then, from the differential equations and the boundary conditions the following behavior follows for the physically relevant quantities:

$$\begin{aligned} m(x)&=(x_1/3^{3/2})-(3x_1{}^3/4)s^2+\cdots\\ Q^2(x)&=1-(3^{3/2}/2)x_1{}^2s^2+\cdots\\ e^{-\lambda}&=(\tfrac{1}{3})+(\tfrac{2}{3})s+(\tfrac{2}{3})[(3^{3/2}x_1{}^2/4)-1]s^2+\cdots \quad (88)\\ e^{\nu}&=(\tfrac{1}{3})+(\tfrac{2}{3})s-(\tfrac{2}{3})s^2+\cdots\\ \omega/4\pi&=-3^{1/2}s+\cdots. \end{aligned}$$

The power series expressions for $e^{-\lambda}$ and e^ν join on smoothly to the accurate exterior values,

$$e^\nu=e^{-\lambda}=(1-2x_1/3^{3/2}x). \quad (89)$$

Near the inner limit, for $x\sim x_{\min}$, write

$$x=x_{\min}(1+u). \quad (90)$$

Then

$$\begin{aligned} m(x)&=(2^{1/2}/3)(x_1{}^4/x_{\min})u^{3/2}+\cdots\\ Q^2(x)&=(x_{\min}{}^2/x_1{}^2)+(2^{3/2}/3)x_1{}^2u^{3/2}+\cdots\\ e^{-\lambda}&=1-(2^{3/2}/3)(x_1{}^4/x_{\min}{}^2)u^{3/2}+\cdots \quad (91)\\ e^{\nu}&=(x_{\min}{}^2/x_1{}^2)+(2^{7/2}/15)x_1{}^2u^{5/2}+\cdots\\ \omega/4\pi&=2^{1/2}u^{1/2}+\cdots \end{aligned}$$

and, for $x_1\leqslant x_{\min}$,

$$e^{-\lambda}=1;\quad e^\nu=(x_{\min}{}^2/x_1{}^2);\quad m=0.$$

It is reasonable to assume a trial value for the eigenvalue parameter x_1, and start a numerical integration working inward from $x_{\max}=x_1/3^{1/2}$ with the starting series (88). The solid angle will first increase, then decrease. When it goes to zero, one wants $e^{-\lambda}$ to be unity. This condition is not satisfied in general. Accordingly, a new choice for x_1 is made. One proceeds by trial and error until $e^{-\lambda}$ goes to 1 as the solid angle goes to zero. This procedure was not adopted because the series expansion (88) was not available when the numerical work was done.

In the procedure that was used the coupled Eqs. (85) and (86) were integrated from the interior boundary $x_{\min}$ although the initial value of Q^2 depends not on a chosen value of x_1 alone, but also on the value of $x_{\min}$. To make the integration depend on a single parameter, the invariance of the equations to scale

change, similar to that used in reference 1, was utilized. Defining $\bar{x}$, $\bar{m}$ and $\bar{Q}^2$ and c by the relations

$$x=b\bar{x},\quad Q^2=b\bar{Q}^2,\quad m=b\bar{m},\quad x_1^2=bc^2; \qquad (92)$$

one has for $\bar{m}$ and $\bar{Q}^2$

$$d\bar{m}/d\bar{x}=(\bar{x}^2/2\bar{Q}^4(1-2\bar{m}/\bar{x})^2(\omega/4\pi)$$
$$d\bar{Q}^2/d\bar{x}=(\bar{x}/\bar{Q}^2(1-2\bar{m}/\bar{x})^3)\{(\omega/4\pi)+(\omega/4\pi)^3/3\},$$

where

$$\omega/4\pi=[1-c^2\bar{x}^{-2}\bar{Q}^2(1-2\bar{m}/\bar{x})]^{\frac{1}{2}}. \qquad (93)$$

The choice $b=x_1^2/\bar{x}_{\min}^2$ or, equivalently, $c=\bar{x}_{\min}$, gives the simple boundary conditions at $\bar{x}_{\min}$:

$$\bar{Q}^2(c)=1,\quad d\bar{Q}^2/d\bar{x}|_c=0,$$
$$\bar{m}(c)=0,\quad d\bar{m}/d\bar{x}|_c=0; \qquad (94)$$

and the integration can be carried through for any c. At the outer boundary, $\bar{x}_{\max}=\Gamma$, the correct boundary values, following from (88) are

$$\bar{Q}^2(\Gamma)=3(\Gamma/c)^2,\quad d\bar{Q}^2/d\bar{x}|_\Gamma=0,$$
$$\bar{m}(\Gamma)=\Gamma/3,\qquad d\bar{m}/d\bar{x}|_\Gamma=0; \qquad (95)$$

but in general these will not be satisfied for a particular choice of c. Starting at an arbitrary c the equations were integrated, on a punched card programed electronic computer, using the Kutta-Runge method; $\bar{m}$ was plotted as a function of $\bar{x}$. If a c is chosen smaller than the actual eigenvalue then, when $\bar{m}(\bar{x})=\bar{x}/3$, the slope $(d\bar{m}/d\bar{x})$ will not vanish. In such cases the integration was stopped at the point where $\bar{m}(\bar{x})=\bar{x}/3$ and the slope examined. A larger c, still less than the eigenvalue, will reduce the slope at the critical point. If c is chosen greater than the eigenvalue then $d\bar{m}/d\bar{x}$ vanishes before $\bar{m}(\bar{x})$ has decreased to $\bar{x}/3$. Thus, the correct eigenvalue for c can be approached both from above and below. With the limited machine time available the nearest value of c to its true eigenvalue for which a numerical integration was carried out was 1.875: the corresponding values for Γ and $\bar{m}(\infty)$ are $\Gamma=4.05$ and $\bar{m}(\infty)=1.35$. In Fig. 2 $m(x)$ and $Q^2(x)$ are plotted, by a dotted line, for this value of c, for which $b=0.071$ (and therefore $x_{\min}=0.134$ and $x_{\max}=0.289$). The smooth curves represent an estimate of $m(x)$, $Q^2(x)$, $e^{-\lambda(x)}$, $e^{\nu(x)}$ and $\omega(x)/4\pi$ for the correct eigenvalue.

5. VARIATIONAL PRINCIPLE FOR THE THERMAL GEON

All the differential equations of the Maxwell-Einstein theory, can be derived from the action principle

$$\delta I=0, \qquad (96)$$

where

$$I=\int\int\int\int \mathfrak{L}(-g)^{\frac{1}{2}}dx^1dx^2dx^3dx^4 \qquad (97)$$

and where

$$16\pi c\mathfrak{L}=c^4/GR_\alpha{}^\alpha-F_{\alpha\beta}F^{\alpha\beta} \qquad (98)$$

and where the independent variables are the ten metric quantities g^{ik} and the four electromagnetic potentials Aj. Similarly, one can derive the differential equations of the simple spherical geon,[1]

$$d^2\phi/dx^2+jk\phi=0$$
$$dk/dx+\phi^2=0 \qquad (99)$$
$$dj/dx=3-[1+(d\phi/dx)^2]/k^2$$

from the variational principle

$$\delta I=0$$
$$I=\int\mathfrak{L}(u)du$$
$$du=k-dx \qquad (100)$$
$$\mathfrak{L}=k^{-1}+3k+jkdk/du+j\phi^2-k-(d\phi/du)^2,$$

as recently shown by Ernst.[2] Encouraged by this result, we found that we could derive the equations for the thermal geon from the principle

$$\delta I=0$$
$$I=\int\mathfrak{L}(x)dx$$
$$\mathfrak{L}(x)=(1-xd\lambda/dx)e^{\nu/2-\lambda/2}-e^{\nu/2+\lambda/2} \qquad (101)$$
$$-(x^2/3)e^{\lambda/2}(e^{-\nu}-x_1^2x^{-2})^{\frac{3}{2}}.$$

However, here the eigenvalue parameter x_1, which we would like to know, already enters the variational principle. The situation resembles that in which we would find ourselves if we tried to express the content of the Schrödinger equation in a variational principle built upon the function

$$\mathfrak{L}(x)=(\nabla\psi)^2+(V-\epsilon)\psi^2. \qquad (102)$$

We will always find the trivial solution $\psi(x)=0$ for this variational problem unless ϵ happens to be an eigenvalue: not a very happy way to find eigenvalues! Much more appropriate is the more familiar Ritz variation principle

$$\epsilon\equiv\frac{\int[(\Delta\psi)^2+V\psi^2]dx}{\int\psi^2dx}=\text{minimum}, \qquad (103)$$

into which we can substitute *any* well-behaved trial function and obtain an upper limit on ϵ.

Correspondingly, we take for variation principle

$$\delta(A^*B^*)=0, \qquad (104)$$

$$A^*=\int_b^1 e^{\lambda^*/2+\nu^*/2}(d/dy)[y(1-e^{-\lambda^*})]dy, \qquad (105)$$

$$B^*=\int_b^1 (y^2/3)e^{\lambda^*/2}(e^{-\nu^*}-3y^{-2})^{\frac{3}{2}}dy. \qquad (106)$$

The independent variable here is y:

$$x=sy;\quad \lambda(x)=\lambda^*(y);\quad \nu(x)=\nu^*(y);$$
$$e^{-\lambda^*}=1-2m^*/y;\quad s=x_{\max}=x_1/3^{\frac{1}{2}}; \qquad (107)$$
$$x_{\min}=sy_{\min}=sb.$$

The boundary conditions are

$$e^{-\lambda^*}=m^*=e^{\nu^*}=\tfrac{1}{3} \quad \text{at} \quad y=1 \tag{108}$$

$$\left.\begin{array}{l} e^{\lambda^*}=1, \quad m^*=0, \quad e^{\nu^*}=b^2/3 \\ d\lambda^*/dy=0, \quad dm^*/dy=0 \end{array}\right\} \text{ at } \quad y=b. \tag{109}$$

From the variational principle (104) follow differential equations for λ^* and ν^* as functions of y which are identical in form to Eqs. (78) and (79) for λ and ν as functions of x *except* for the introduction on the right-hand side of (78) and (79) of an extra factor

$$s^2=A^*/B^*. \tag{110}$$

The logic proceeds so: (1) Introduce into (105), (106) trial functions $\lambda^*(y)$ and $\nu^*(y)$ that satisfy (108), (109) and that contain one or more adjustable parameters, among them the lower limit b. (2) Calculate (A^*B^*) and extremize with respect to choice of these parameters. (3) From (110) calculate s^2 and hence x_1 and the geon mass

$$M=M_T x_1/3^{\frac{1}{2}}=M_T s/3. \tag{111}$$

If instead the quantities λ and ν are calculated as functions of x by solution of *differential* equations (78) and (79) for a trial eigenvalue x_1, an improved estimate of the eigenvalue is given by

$$x_1/3^{\frac{1}{2}}=s=(AB)^{\frac{1}{2}},$$

where A and B are the analogs of (105) and (106) for $\lambda(x)$ and $\nu(x)$.

6. PHOTON-PHOTON COLLISIONS IN A THERMAL GEON

The electromagnetic energy content of a thermal geon decreases slowly not only by the monomolecular process of barrier penetration but also by bimolecular processes in which two photons collide, either to produce a pair of electrons, or to go off as photons in new directions.[1] We wish to estimate the rate of these two processes in a very active part of a thermal geon: at the radius of the stable circular orbit (Fig. 1), where the solid angle occupied by bound rays is the largest,

$$(\text{solid angle}/4\pi)\doteqdot 0.59 \tag{112}$$

and where the effective temperature is

$$T_{\text{eff}}=Te^{-\nu/2}\doteqdot 3.3T. \tag{113}$$

Here the number of photons Ω_1 per unit volume in the interval of circular wave number $d\Omega_1$, and in the interval of solid angle $d\omega$ (within the allowed cone) will be

$$dn_1=(d\omega/4\pi)\Omega_1^2 d\Omega_1[-1+\exp(\hbar c\Omega_1/T_{\text{eff}})]^{-1}. \tag{114}$$

The collision between two photons[10] of wave numbers Ω_1 and Ω_2 whose directions of motion make an angle θ will look like the collision between two photons of equal wave number Ω^* and opposite direction, in a suitably selected local Lorentz system, where

$$\Omega^*=(\Omega_1\Omega_2)^{\frac{1}{2}}\sin\tfrac{1}{2}\theta. \tag{115}$$

The total collision cross section, integrated over all angles of the emergence of the pair (for process 1) or of the scattered photons (for process 2) will be

$$\sigma=\sin^2\tfrac{1}{2}\theta\sigma^*(\Omega^*), \tag{116}$$

where Ω^* is the cross section calculated for equally energetic but oppositely moving photons. For the pair process the cross section σ^* vanishes below the threshold, $\Omega^*=mc/\hbar$, and reaches a maximum value of the order (e^2/mc^2) for a wave number that is a small multiple of this threshold value. For temperatures small compared to mc^2 it follows from the Planck formula that the number of pair production processes is exponentially small, with an exponential factor qualitatively of the form

$$\exp-(\mu mc^2/T), \tag{117}$$

where μ is of the order of unity.

The cross section for elastic photon-photon collisions also reaches a peak for wave numbers, Ω^*, of the order $mc/\hbar$. The peak value of σ^* for this process is of the order $(e^2/\hbar c)^2(e^2/mc^2)^2$, much smaller than that for the pair production mechanism. However, the cross section has no threshold and varies at low wave numbers in accordance with the formula[11]

$$\sigma^*=(52/1125\pi)(e^2/\hbar c)^2(e^2/mc^2)^2(\Omega^*\hbar/mc)^6. \tag{118}$$

For low temperatures the effective cross section for collision between two photons consequently varies as T^6, and therefore dominates over an exponentially small factor of the form (117). For this reason we can disregard pair production processes, relative to elastic collisions, so long as the temperature is considerably less than mc^2.

Not all photon collision processes result in loss of photons from the system. In some cases one or both of the new quanta still move in bound orbits. Only those are lost whose directions are thrown outside the cone of trapping angles. In particular a new photon escapes from the system if it is moving along a radius.

For an order-of-magnitude estimate of the rate of loss of energy we disregard details of the differences between trapping and escaping directions and consider the product of the following factors:

number of photons Ω_1 per cm^3	$\sim T^3/\hbar^3c^3$
number of photons Ω_2 per cm^3	$\sim T^3/\hbar^3c^3$
collision velocity	c
cross section	$\sim(e^2/\hbar c)^2(e^2/mc^2)^2(T/mc^2)^6$
energy loss from geon on collision	$\sim T$

(119)

product, energy loss per cm^3 and per sec $\sim(e^2/\hbar c)^4(mc^2/\hbar)\times(mc^2)^4\hbar^{-3}c^{-3}(T/mc^2)^{13}$

[10] G. Breit and J. A. Wheeler, Phys. Rev. 46, 1087 (1934).

[11] H. Euler and B. Kockel, Naturwissenschaften 23, 246 (1935); W. Heisenberg and H. Euler, Z. Physik 98, 714 (1936).

In contrast, the energy on hand per unit volume is of the order $T^4/\hbar^3c^3$. This quantity, divided by (119), fixes a characteristic scale of time, τ, for depletion of the geon by photon-photon collisions:

$$\begin{aligned}\tau^{-1} &\sim \frac{\text{(energy loss/cm}^3\text{ sec)}}{\text{(energy/cm}^3\text{)}} \\ &\sim (e^2/\hbar c)^4(mc^2/\hbar)(T/mc^2)^9 \\ &\sim (10^{12}\text{ sec}^{-1})(T/mc^2)^9.\end{aligned} \tag{120}$$

For a characteristic time as long as a year it is sufficient to have a temperature of the order

$$T\sim 10^{-20/9}mc^2\sim 10^{-2}mc^2$$

corresponding to a geon mass of the order

$$M_T\sim 10^{44}g$$

and radius

$$R_T\sim 10^{16}\text{ cm}.$$

As the energy loss continues, the thermal geon shrinsk, grows denser and hotter, and loses energy at a rapidly increasing rate. As the temperature rises to the neighborhood of mc^2, pair production processes rapidly increase in importance. Then the physics of the system takes on quite a different character which we do not analyze here.

7. ZERO-POINT ENERGY

We have consistently disregarded quantum effects, or rather have consistently attempted to choose conditions where quantum effects are unimportant, in all except the considerations of very hot geons in the last section. However, at all temperatures one has to reckon with zero-point fluctuations in the electromagnetic field, as well as with the fluctuations due to the thermal radiation itself. One can formally associate these zero-point fluctuations with a zero-point energy, $\frac{1}{2}\hbar c\Omega$, that goes with each field oscillator. Usually this energy is left out in the bookkeeping of the energy of the electromagnetic field. The subtracted density of zero point energy of the vacuum is infinite. Normally one deals with field physics at the level of special relativity, where such an infinite quantity can be disregarded. However, in general relativity there is no such thing as an arbitrary additive constant in the density of field energy. Energy density curves space, and an increase in energy density produces an increase of curvature. Our analysis of thermal geons is based on the tacit assumption that the zero-point energy does not have to be counted, either as energy or as a source of curvature. We hope this point of view is correct, as it gives reasonable results in familiar situations. Nevertheless, a deeper approach to the problem of zero-point energy is needed—a problem whose overriding importance to all of field physics Niels Bohr has often stressed.

APPENDIX ON RAY-WAVE EQUIVALENCE

There is a close correlation between the field theory solutions of (44), (45), and (50) and the motion of a classical corpuscle of light or "pseudophoton" along a geodesic in the same gravitational field. Superposition of standing waves of slightly different Ω_{nl}, l, and m values allows one to build up a wave packet. This concentration of energy will remain the better defined in space and time the larger are the relevant values of l and n—that is, the shorter the wavelength of the disturbances of significant amplitude in comparison with the scale of distances over which the gravitational field changes appreciably. The correspondence between waves and pseudophoton orbits in the idealized limit which disregards the spreading of a wave packet may be analyzed this way: (1) We pass from the wave itself to its phase,

$$\text{(vector potential)} = \left\{\begin{matrix}\text{slowly varying}\\ \text{vector function}\\ \text{of position}\end{matrix}\right\} e^{i\cdot\text{phase}} \tag{A1}$$

where $e^{i\cdot\text{phase}}$ varies rapidly with position; (2) we approximate the differential equation for the phase by neglecting the slow change of the amplitude with distance in comparison with the rapid change of the phase with distance. Thus we pass from the accurate phase of (A1) to the pseudophase (or eikonal)—the central concept in William Rowan Hamilton's method of treating problems in geometrical optics as well as in mechanics.* The equation for the pseudophase has the Hamilton-Jacobi form

$$g^{\alpha\beta}(\partial\Phi/\partial x^\alpha)(\partial\Phi/\partial x^\beta)=0. \tag{A2}$$

[Instead of deriving this from the wave equation by the substitution (A1), one can start directly with the picture of a surface propagating parallel to itself with the speed of light. The surface and its propagation with time can be expressed in the form

$$\Phi(x^1,\cdots x^4)=17.2. \tag{A3}$$

Let x^4 be one value of the cotime, and x^1, x^2, x^3 the coordinates of one point on the surface at that time. Let x^i+dx^i denote a neighboring point on the same moving surface, $\Phi=17.2$, at a slightly later time. The space and time separations of the two points are connected by the relation

$$(\partial\Phi/\partial x^\mu)dx^\mu=0. \tag{A4}$$

The requirement that the surfaces of (A4) move with the speed of light leads directly to (A2).] (3) We recognize not only single solutions, $\Phi(x_1'\cdots x^4)$ of the pseudophase equation, but also a typical family of

---

* See, for example, J. L. Synge, *Geometrical Mechanics and de Broglie Waves* (Cambridge University Press, 1954), Chap. II; or L. Landau and E. Lifschitz, *The Classical Theory of Fields*, reference 8, p. 271.

pseudophase functions,

$$\Phi = \text{function of } (x^1, \cdots x^4; \beta^1, \beta^2, \beta^3) + \beta^4(\beta^1, \beta^2, \beta^3), \quad \text{(A5)}$$

where each number of the family is characterized by a specific set of values for the three numbers β^1, β^2, β^2. These numbers might correspond in the case of physical optics to components of the wave vector for a monochromatic wave, but need not have such an immediate interpretation. The fourth constant of integration β^4, represents an additive phase constant, always allowed because (A2) contains only derivatives of the pseudophase, never Φ itself. (4) We impose the requirement of constructive interference. Several neighboring solutions of the wave equation of the form (A1) are superposed in such a way as to build up a wave packet. To make this requirement clean cut, we go to the pseudophase picture, and enquire how a pseudophoton must move in order that it shall always lie at the point where four pseudophase functions of nearly identical β values have a common value. The pseudophase need not be constant along the path of the pseudophoton. It is only required that at each point of the path, $x^i(\sigma)$, the four pseudophase functions should have a common value:

$$\begin{aligned}
&\Phi(x(\sigma);\ \beta^1,\beta^2,\beta^3;\ \beta^4(\beta^1,\beta^2,\beta^3))\\
&=\Phi(x(\sigma);\ \beta^1+d\beta^1,\beta^2,\beta^3;\ \beta^4(\beta^1+d\beta^1,\beta^2,\beta^3))\\
&=\Phi(x(\sigma);\ \beta^1,\beta^2+d\beta^2,\beta^3;\ \beta^4(\beta^1,\beta^2+d\beta^2,\beta^3))\\
&=\Phi(x(\sigma);\ \beta^1,\beta^2,\beta^3+d\beta^3;\ \beta^4(\beta^1,\beta^2,\beta^3+d\beta^3))
\end{aligned}$$

or

$$\partial\Phi(x;\beta)/\partial\beta^s + (\partial\Phi/\partial\beta^4)(\partial\beta^4/\partial\beta^s) = 0 \ (=1,2,3). \quad \text{(A6)}$$

These three equations connect the three position coordinates of the pseudophoton with the time, and therefore suffice completely to determine its motion. From (A2) and its first derivatives with respect to the β's, this motion proceeds along a null geodesic,

$$d^2x^k/d\sigma^2 + \Gamma_{\mu\nu}{}^k (dx^\mu/d\sigma)(dx^\nu/d\sigma) - \begin{pmatrix}\text{undetermined}\\ \text{function of } \sigma\end{pmatrix} \cdot (dx^k/d\sigma) = 0$$

as expected. As an illustration how one gets all details of the pseudophoton motion from (A6), one can ask for the velocity components of the motion. For this one differentiates (A6) once with respect to σ and solves for the $dx^i/d\sigma$ by the method of determinants, finding

$$dx^i/d\sigma = f(\sigma)\{i\lambda\mu\nu\}(\partial^2\Phi/\partial x^\lambda\partial\beta^1)(\partial^2\Phi/\partial x^\mu\partial\beta^2) \times (\partial^2\Phi/\partial x^\lambda\partial\beta^3). \quad \text{(A7)}$$

Here the arbitrary functions $f(\sigma)$ in the solution appear because σ itself represented an arbitrary parametrization of the path of the pseudophoton. The same method gives the acceleration and other details of the motion.

For the spherically symmetric metric of the thermal geon the pseudophase propagation equation takes the form

$$e^{-\lambda}(\partial\Phi/\partial r)^2 + r^{-2}(\partial\Phi/\partial\theta)^2 + (r\sin\theta)^{-2}(\partial\Phi/\partial\phi)^2 - e^{-\nu}(\partial\Phi/\partial T)^2 = 0 \quad \text{(A8)}$$

which possesses separable solutions of the form

$$\Phi = \int_{r\min}^{r} [1 - l(l+1)e^\nu/\Omega^2 r^2]^{\frac{1}{2}} \Omega e^{\lambda/2-\nu/2} dr + \int_{\theta\min}^{\theta} [l(l+1) - m^2 \sin^2\theta]^{\frac{1}{2}} d\theta + m\phi - \Omega T + \beta^4(\beta^1,\beta^2,\beta^3) \quad \text{(A9)}$$

Here the three constants of integration, β^1, β^2, β^3, have been expressed in the form Ω, l, and m to bring out the identity between the pseudophase and the phase of the JWKB approximation to the solution of the wave equation. We can superpose waves of different l, m, and Ω values to build up a wave packet that will trace out the designated geodesic.

The nature of the geodesic curve is independent of its orientation; consequently it is sufficient to consider m values close to zero, corresponding to motion in a meridian plane. *Which* meridian can be specified in the wave picture by the relative phase or β^4 values with which one superposes waves of slightly different m values; in the eikonal formulation, by setting equal to zero the derivative of the pseudophase with respect to m; thus $\partial\Phi/\partial\beta^3 = \partial\Phi/\partial m = 0$ gives an equation for ϕ. However, as there is no interest in this angle, it is appropriate to overlook this relation, and set m equal to a fixed value: $m=0$. Likewise the first equation of stationary pseudophase, $\partial\Phi/\partial\beta^1 = \partial\Phi/\partial\Omega = 0$, is also irrelevant for our purpose; we do not care *when* the pseudophoton arrives at a given point in its orbit. We are left with the second equation of constructive interference, $\partial\Phi/\partial\beta^2 = \partial\Phi/\partial l = 0$, to determine the *shape* of the geodesic:

$$0 = -\int_{r\min}^{r} [r^2/e^\nu P^2 - 1]^{-\frac{1}{2}} e^{\lambda/2} dr/r + \theta + \text{const.} \quad \text{(A10)}$$

Here the quantity P is an abbreviation for the expression $[l(l+1)]^{\frac{1}{2}}/\Omega$ and represents the *impact parameter* of the pseudophoton—the distance of closest approach in the absence of gravitational forces. Evidently *this single constant determines the shape of the geodesic*, not Ω or l individually. From (A10) we derive the properties of the geodesics already discussed in the main text and pictured in Fig. 1. All the information gained about geodesics carries over to the characteristic solutions of the wave equation. The limits of motion, $r_1(P)$ and $r_2(P)$, of the rays correspond to the points where the field amplitudes change from oscillation to exponential fall off. Inclination of the nodal surfaces of the field is closely related to the inclination of the rays; and the energy carried by the rays or the wave fields has to supply the gravitating mass that holds the geon together.

Linear and Toroidal Geons

FREDERICK J. ERNST, JR.*

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

ONE particular phase of Wheeler's program of drawing a new richness from general relativity involves the concept of "mass without mass." The bodies, which Wheeler constructs solely from gravitational and electromagnetic fields, have a mass which is due to the electromagnetic field energy alone. No more concrete form of matter is present in these geons.

For an undergraduate thesis topic, I decided to investigate certain types of geons, in particular those in which the electromagnetic field energy is spread over an infinitely long line, and those in which it is concentrated in a toroidal region of space. The former seemed to permit simple analysis, while the latter seemed to be the type most fundamental in geon theory.

The problem of describing the gravitational and electromagnetic fields of a linear geon was treated in the self-consistent field approximation, by solving Maxwell's equations in the *average* gravitational field produced by the electromagnetic energy of the geon.[1] For a linear geon whose mass-per-unit-length is M and whose electromagnetic field energy is concentrated in a cylinder of radius r, the following relation was found for the product of the frequency and the reduced wavelength of the electromagnetic waves which travel along the line of energy concentration:

$$\Omega\lambda\!\!\!^{-}/c \cong (r/a)^N, \quad N = 8(GM/c^2) - 8(GM/c^2)^2.$$

Here G is the gravitational constant, c the velocity of light, and a a constant with dimensions of length. This constant a is undetermined for a geon which is really infinitely long. However, if the linear geon is bent into a large circle of radius $R \gg r$, the boundary conditions at infinity fix the value of a at approximately $8R$.

Probably the greatest usefulness of the geon concept lies in the fact that it constitutes a completely consistent, completely covariant, picture of an extended body; the problem of treating internal stresses consistently never arises. No one believes that they bear the slightest relationship to any existing bodies or elementary particles.

---

* Now at University of Wisconsin, Madison, Wisconsin.

[1] Frederick J. Ernst, Jr., Phys. Rev. **105**, 1665 (1957).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Feynman Quantization of General Relativity*†

CHARLES W. MISNER

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

1. STRUCTURE OF QUANTUM THEORY OF GENERAL RELATIVITY

THIS paper reports the beginnings of the quantum theory of general relativity based on the Feynman[1,2] integral or "sum over histories." We have not seen a way to use the Feynman integral to solve immediately all the principal problems. We have to study the theory one piece at a time and to set each fragment in place when we are able to understand it. In this sort of approach we need not follow any logical order, but may study the easy parts first and hope to fill in the rest later. However, some over-all picture of what the completed puzzle may look like is necessary in order to recognize the pieces.‡

This paper describes in broad outline the principal features of a quantum theory of general relativity and fills in a few details. If rigor could be supplied we would have a theory rather than an approach to one.

Two fragmentary but concrete contributions to the theory based on the Feynman integral are made. (1) We formulate an *H principle* (Sec. 4), which specifies the relative weights to be given to different values of the fields in the Feynman integral. This principle—applied to the metrics of general relativity (Sec. 5)—leads directly to an operator form of the field equations (Sec. 6). (2) We make (Sec. 7) a partial evaluation of the Feynman propagator. In consequence we are able to write down immediately the state of the field on any of a rather wide class of hypersurfaces when we have specified the state on one hypersurface. This evaluation is possible because the answer is trivial. The state is essentially identical on all hypersurfaces of the same class. In other words, we prove the important result that *the Hamiltonian operator is zero*. This situation is peculiar to a theory in which the metric is quantized.

---

* A part of this work (Sec. 5) was done while the author held a Charlotte Elizabeth Proctor Fellowship (1954–1955) which he gratefully acknowledges.

† Based in part on a thesis submitted to Princeton University, May 1957, in partial fulfillment of the requirements for Ph.D.

[1] R. P. Feynman, thesis, Princeton University, 1942.

[2] R. P. Feynman, Revs. Modern Phys. **20**, 367 (1948).

‡ *Note added in proof.*—Two other papers meant to be read in association with this paper, C. W. Misner and J. A. Wheeler, "Classical Physics as Geometry," and J. A. Wheeler, "On the Nature of Quantum Geometrodynamics," were completed too late for publication in this issue and have been submitted for publication elsewhere.

2. ALTERNATIVE APPROACHES TO THE QUANTIZATION OF GENERAL RELATIVITY

Four approaches have been suggested to discover the content of the quantum theory of general relativity: (1) One considers an ideal Lorentz space, and treats the departures of the actual metric from this flat condition as small—that is, one linearizes the gravitational field equations, applies standard methods of field quantization,[3,4] and then attempts to modify this basically linear theory by reinstating the nonlinear terms.[5,6] All other approaches consider the field equations of general relativity in their fully covariant but nonlinear form. (2) The canonical formalism endeavors to investigate the noncommutative algebra of the operators of the theory in general and the Hamiltonian operator in particular, with a view to finding ultimately in this way eigenvalues and transition probabilities.[7–9] One first sets up an appropriate algebraic structure on the field variables by defining Poisson or Dirac[10] brackets. Here "appropriate" is an abbreviation for two conditions: (a) the algebraic formulation of the classical theory is equivalent to the Lagrangian formulation, and (b) the algebraic structure admits a representation by linear operators in Hilbert space where the brackets are represented by commutators. The problem thus defined has next to be *solved* by finding a particular operator representation.

The other two approaches deal with propagators, $(f_2\sigma_2|f_1\sigma_1)$, that are natural generalizations of the propagator,

$$\langle x_2t_2|x_1t_1\rangle=[m/2\pi i(t_2-t_1)]^{\frac{1}{2}} \times \exp im(x_2-x_1)^2/2(t_2-t_1) \quad (1)$$

of a simple one particle problem.

The hypersurfaces, σ_2 and σ_1, take the place of the time variables, t_2 and t_1; and the field configurations on these surfaces, f_2 and f_1, take the place of the position coordinates, x_2 and x_1. (3) The action principle

---

[3] L. Rosenfeld, Ann. Physik **5**, 113 (1930).

[4] W. Pauli and M. Fierz, Helv. Phys. Acta **12**, 297 (1939).

[5] A summary of such linearized quantum analyses has been given by F. Belinfante, Revista Mex. Fis. **4**, 192 (1955).

[6] B. E. Laurent, Nuovo cimento (10) **4**, 1445 (1956).

[7] This approach has been intensively studied by P. Bergmann Revs. Modern Phys. **29**, 352 (1957), Helv. Phys. Acta., Suppl. IV, 79–97 (1956).

[8] Belinfante, Caplan, and Kennedy, Revs. Modern Phys. **29**, 518 (1957).

[9] B. S. DeWitt, Revs. Modern Phys. **29**, 377 (1957).

[10] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); **3**, 1 (1951).

of Schwinger[11] deals with the infinitesimal changes in this propagator,

$$\delta\langle f_2\sigma_2|f_1\sigma_1\rangle=(+i/\hbar)\langle f_2\sigma_2|\delta I|f_1\sigma_1\rangle, \quad (2)$$

which come about through changes in the action, I' either by way of alterations in the position of the two surfaces, in the configurations of the fields on the surfaces, or otherwise. In distinction to the canonical formalism with its use of Hamiltonian and momenta, Schwinger's method deals with the action and expresses itself in a manifestly covariant form. This method has not been applied independently to general relativity, but is used in conjunction with the Feynman method. (4) The Feynman method focuses, not on a differential equation to be solved for the propagator, but on a formula for the solution:

$$(f_2\sigma_2|f_1\sigma_1)=N^{-1}\sum_H \exp(iI_H). \quad (3)$$

With this expression goes a formula for a matrix element which Dyson[12] writes in the form,

$$(f_2\sigma_2|\mathcal{O}|f_1\sigma_1)=N^{-1}\sum_H \mathcal{O}_H \exp(iI_H). \quad (4)$$

Here H indicates a field history, that is, a definite specification of the field throughout the region between the hypersurfaces σ_2 and σ_1. The quantity $\mathcal{O}$ is a functional of the field and $\mathcal{O}_H$ is its value for the particular field history H. The sum extends only over field histories whose boundary values on σ_2 and σ_1 are f_2 and f_1. The normalization factor N depends on σ_2 and σ_1, not on f_2 and f_1, and is introduced to secure the unitarity of the propagator. The (dimensionless) action is taken in general relativity to have the value

$$I_H=(c^3/16\pi G\hbar)\int R(-g)^{\frac{1}{2}}d^4x. \quad (5)$$

The Newtonian gravitation constant, G, and the quantum of angular momentum, $\hbar$, appear in the theory, never individually, but only in a combination with the dimensions of length,

$$L^*=(\hbar G/c^3)^{\frac{1}{2}}=1.62\times10^{-33}\text{ cm}. \quad (6)$$

Like the Schwinger method, the Feynman method deals with manifestly covariant quantities. The operators that are associated with physical quantities are *defined* by integrals of the type (4), where every quantity on on the right-hand side of the equation is a c number. Another convenient quality of the Feynman method is that it is flexible and can be applied in a limited way. An explicit representation of an algebraa dequate to express the entire classical theory—such as the canonical method demands—contains a vast and indigestible amount of information. In the Feynman theory we need not consider all these operators at once; we may begin by constructing just one of them, or some other operator which we think might be simpler than the basic operators of the canonical theory. In this way the theory may be attacked in a succession of short skirmishes rather than in a single frontal assault.

What of the problems of the three methods (1), (2), and (4) that have received some detailed consideration? We discuss here (1) the inappropriateness of the linearized treatment, (2) the problem of observables in the true theory of general relativity, and its different immediate consequences for the (a) canonical and (b) Feynman type of quantization, and (3) the work up to now on the methodology of Feynman quantization.

In a linearized version of general relativity it is easy to make calculations by standard perturbation methods. However, the characteristic length L^* is so small compared to the distances relevant in any familiar experimental context that such computations are not of much interest. If gravity is to occupy a significant place in modern physics, it can do so only by being *qualitatively* different from other fields. As soon as we assume gravity behaves qualitatively like other fields, we find that it is quantitatively insignificant.

Classical theory contains no characteristic length L^*. There we do have the possibility to confine our attention —if we wish—to the realm of weak fields. Consider the classical field equations,

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=0, \quad (7)$$

and an initial space-like surface, σ. On it admit as initial condition only such a metric as leaves σ nearly flat. Moreover, permit only small normal derivatives of the metric at σ. Then the same weak field conditions will prevail some time later. Therefore the linear approximation makes sense.

Not so in quantum theory. We can form a state which makes the initial surface σ nearly flat, even on a submicroscopic level. However, uncertainty relationships will be expected to prevent our simultaneously restricting the time derivative of the metric at σ to a small value. Consequently large curvatures over small regions are to be expected. Similarly, even if there were a "vacuum state" for the gravitational field, there is no reason to expect that it would be a state where the metric is nearly flat over *small* regions—a point that has been emphasized by Professor Wheeler. In other words, a linearized version of general relativity seems precisely adapted to throwing out just those features which are physically new and interesting. We therefore pass it by.

In the full nonlinear theory one meets the fundamental issue, what are the real physical observables? We always use the word observable in a classical sense: an *observable* $\mathcal{O}$ is a functional of field histories H for

[11] J. Schwinger, Phys. Rev. 82, 914 (1951).
[12] F. J. Dyson, *Advanced Quantum Mechanics* (Cornell University, Ithaca, 1954, mimeographed), p. 54.

which there is in classical theory a conceptual experiment whose outcome is interpreted as providing the value $\mathcal{O}_H$ of $\mathcal{O}$ for the unique field history H which existed physically at the performance of the experiment. This definition of an observable avoids using the idea of an *event*, but the simplest examples of observables require this idea. In electromagnetic theory, for instance, where the field in question is the vector potential A, an event ξ determines a functional $\mathcal{A}_\xi$ whose value for a field history A is $\mathcal{A}_\xi(A)=A(x_\xi)$. Here x_ξ is the point corresponding to the event ξ. This functional, $\mathcal{A}_\xi$, is not an observable; however, $\mathcal{F}_\xi(A)=F(x_\xi)=\text{curl}A(x_\xi)$ defines an observable functional $\mathcal{F}_\xi$. In electromagnetic theory it is usual to speak of $F(x)$ as an observable when one really means the observable $\mathcal{F}_\xi$. The fancy $\mathcal{F}_\xi$ is in fact superfluous since every physicist knows what is and is not observable (ideally) in classical electromagnetic theory. There is nothing to be gained there by explicitly distinguishing all distinguishable ideas. In general relativity it is better to make the necessary distinctions explicitly.

A fundamental distinction is that between a point x and an event ξ. The possibility of making distinctions is the fundamental fact which makes logical thinking possible. A *mathematical object* is anything which is subject to logical discussion; as such it must satisfy one axiom—that it be distinct from every other mathematical object. A *point* is a mathematical object which is an element of a set $\mathfrak{M}$ satisfying further axioms; for present purposes we require $\mathfrak{M}$ to be a four-dimensional differentiable manifold.[13] These axioms completely exhaust the meaning of the word "point" as we use it.

In studying electromagnetic theory the distinction between points and events is not interesting since we may assume that there is a unique point x_ξ in the theory corresponding to each physical event ξ. The event may be thought of as a time and place where two neutral point particles collide, or where it could have been arranged that such a collision occur. The idea of an event, then, need not involve electromagnetism, and it may be assumed that the correspondence $\xi\leftrightarrow x_\xi$ has been established on the basis of a nonelectromagnetic theory. Thus, before we begin a study of electromagnetism, we already know that a point is a good mathematical model of an event.[14] The correspondence $\xi\leftrightarrow x_\xi$ may be called *the point theory of events*, and on the basis of this theory we may use the events ξ themselves as points in discussing electromagnetic theory.

In general relativity, a different situation prevails. Before studying general relativity we have no notion of a metric, no theory of distance and time. We are therefore unable to imagine any idealized physical theory which would provide a theory of events to use in discussing general relativity. The theory of events must spring up within general relativity, not logically precede it. When a scalar such as $R(x)$ is considered as an observable in classical general relativity, the point x represents an event and is in reality a complicated functional of other events which the observer used to define his location, and of the metric throughout a region containing both those reference events and the event associated with x. If we were prepared to discuss such an observable, we would use a notation that gave fair warning of the complexity of the computation it envisages, and of the large number of quantities on which it depends. When we write $R(x)$ we mean a comparatively simple mathematical object, the value of the curvature scalar at a point x for the metric under consideration. The point is not an observable in the classical theory. Consequently, $R(x)$ is not an observable functional of the metric, nor is $g_{\mu\nu}(x)$, nor is the value of *any* scalar or tensor function at x. If ϕ is any function of points defined by the metric then we expect no corresponding operator $\phi(x)$ to be constructible in the quantum theory of general relativity. For trivial functions like $\phi=0$ we can find corresponding trivial operators. One advantage of the Feynman approach is that it allows us to work with functions like $R(x)$ without having to assume the existence of a corresponding operator.

Bergmann[15] finds that to carry through the canonical quantization it may be necessary to find the "true observables" in general relativity and use them in place of more familiar field variables. Similarly, we expect the Feynman method will generally not provide constructions for operators corresponding to classical quantities which are not observables. As discussed earlier in this section, however, a start can be made in the Feynman theory with only one or two such observables in hand, while the canonical theory seems to require that a large number of them be expressed in manageable form.

Distinct from the problem of characterizing and finding observables (which we have defined as a classical problem) is the problem of describing the *measurability* of such observables in the light of quantum theory. This problem has been discussed on the basis of quantum limitations on the measuring instruments by Osborne,[16] by Anderson,[17] and by Saleker and Wigner.[18] In electromagnetic theory a satisfactory discussion from this point of view has been given by

[13] Georges de Rham, *Variétés Différentiables* (Hermann et Cie, Paris, 1955), p. 1; H. Whitney, Ann. Math. 37, 645 (1936).

[14] Convincing arguments have been given which indicate that a point is not a good mathematical model for a quantum mechanical "event." See E. P. Wigner, Revs. Modern Phys. 29, 255 (1957).

[15] P. G. Bergmann, Nuovo cimento (10) 3, 1177 (1956); see also reference 7.

[16] M. F. M. Osborne, Phys. Rev. 75, 1579 (1949).

[17] J. L. Anderson, Revista Mex. Fis. 3, 176 (1954).

[18] H. Saleker and E. P. Wigner (to be published). See also reference 14.

Bohr and Rosenfeld,[19] but only after an analysis of the measurability had been provided by the quantum theory of the electromagnetic field. The two approaches —one from quantum theory of the measuring apparatus, the second from the quantum theory of the field—are complementary but should be consistent. A discussion of measurability based on the quantum field theory alone is more direct, since it requires no ingenuity in discovering an optimal experimental arrangement. A desire to understand the quantum limitations on measurements of the gravitational field is therefore one reason for investigating the quantum theory of general relativity.

The Feynman method is the basis of our approach to the quantum theory of general relativity, so we now consider the techniques which have been used to express the Feyman method mathematically. Feynman's technique[1,2] in defining the "sum over histories" was based on a use of Hamilton's principal function S. In the quantum mechanics of a particle, the propagator is

$$\langle x''t''|x't'\rangle=\int \exp\left\{\frac{i}{\hbar}\int_{t'}^{t''} Ldt\right\}\delta x, \qquad (8)$$

where L is the Lagrangian. To define this expression Feynman first considered the case of a small time interval Δt, and set

$$\langle x'', t+\Delta t|x't\rangle=A^{-1}\exp\{iS(x'', t+\Delta t; x',t)/\hbar\}, \qquad (9)$$

where S is $\int Ldt$ evaluated for the classical path with the required end points,[20] and A is a normalization factor. An iteration of the infinitesimal propagator (9) together with a limit $\Delta t\rightarrow 0$ produces the "sum over paths" of Eq. (8). The normalization factor A is fixed by requirements inherent in this limiting process; it may depend on x', x'', t, and Δt, but is independent of x' and x'' if the coefficients of the $\dot{x}$ terms in L are independent of x, and is independent of t if L is not explicitly a function of t. This S method of defining the sum over paths has been applied by Choquard[21] to situations where the potential is anharmonic. It has been extended[22] to Lagrangians where the coefficients of the velocity terms are not constants. Anderson[23] has pointed out, however, that this S technique when applied to gauge-invariant theories such as general relativity leads to an unacceptable result: $A^{-1}=0$. The explanation of this failure of the method is easy to find: In a gauge-invariant theory there are infinitely many histories which describe the same physical situation; and if one attempts to normalize after integrating, the result is an unsatisfactory $A^{-1}=0$ normalization. This difficulty has been eliminated in electromagnetic theory by Wheeler[24] and by Laurent[6] who perform an average over gauges in place of an integration. There were, however, two problems which the factor A was to solve. One problem was to determine the absolute weight to be given to a field history in the sum over histories; it was this problem which led to the $A^{-1}=0$ failure, but which is solved by using an average over gauges. The second problem, which logically precedes the first, was to determine, through the dependence of A on x, the relative weights to be given different histories. In Sec. 3 we formulate the H principle which states the relative weights which are to be assigned different field histories. It is then meaningful to speak of an average so that a way is open to consider subsequently the normalization of the "sum over histories." The averaging process, and the remaining, field independent, normalization factor, are not investigated.

More recent treatments of Feynman quantization may be called non-S methods. Most of these non-S treatments are based on the work of Davison[25] who introduced the technique of Fourier analyzing the history and then integrating over the Fourier coefficients. This technique has been applied by Burton and De Borde[26] to several mechanical systems, and is used by Matthews and Salam,[27] and by Laurent[6] in field theory. Laurent's work is the first application of the Feynman method to general relativity. None of these techniques can be applied to a quantization of the full theory of general relativity, since Fourier analysis is applicable only to linear systems.

Wheeler[24] has used an essentially non-S technique to quantize the electromagnetic field. In common with the original Feynman approach, he approximates the space-time volume V between two hypersurfaces by a grid of points $\mathfrak{N}$. A field history is then approximated by a *skeleton history* where the value of the vector potential is prescribed only at the grid points. Next the action integral is also *skeletonized* in being approximated by a sum which is defined for every skeleton history. Finally the propagator

$$(A_{2}\sigma_{2}|A_{1}\sigma_{1})=\int\exp\left\{(-i/16\pi\hbar c)\int F_{\mu\nu}F^{\mu\nu}d^4x\right\}\delta A \qquad (10)$$

is defined by integrating over the values of A at the grid points $\Pi_{x\epsilon\mathfrak{N}}d^4A_x$ and passing to a limit as the mesh

[19] N. Bohr and L. Rosenfeld, Kgl. Danske. Viden selskab, Mat.-fys. Medd. 12, No. 8 (1933).

[20] See for instance, H. Goldstein, *Classical Mechanics* (Addison-Wesley Press, Cambridge, 1950), p. 276.

[21] Philippe Choquard, thesis (École Polytechnique Fédérale, Zurich, 1955).

[22] Cécile Morette, Phys. Rev. 81, 848 (1951).

[23] J. L. Anderson (private communication).

[24] J. A. Wheeler, *Fields and Particles* (unpublished). Lectures given at Princeton (1954–1955 and 1956–1957) and at Leiden (1956).

[25] B. Davison, Proc. Roy. Soc. (London) A225, 252 (1954).

[26] W. K. Burton and A. H. De Borde, Nuovo cimento (10) 2, 197 (1955).

[27] P. T. Matthews and A. Salam, Nuovo cimento (10) 2, 120 (1955).

of the grid tends to zero. The entire procedure may be abbreviated by the phrase

$$\delta A = N^{-1}(V) \prod_{x\epsilon V} d^4 A(x), \tag{11}$$

where the normalization factor N indicates both a finite normalization factor $A(\mathfrak{M})$ which appears with each skeleton integral, and the procedure of averaging over gauges.

Polkinghorne[28] has shown how operator field equations and canonical commutation rules may be derived in a non-S Feynman field theory. He does not require any details of the construction of the Feynman integral, and his work provides a justification for the use of non-S definitions of the functional integral.

In applications of the Feynman method to field theory considered by the British group, the states are treated abstractly and denoted by Dirac kets[26–28]: $|f'\sigma'\rangle$ is a state for which the field f on σ' is characterized by eigenvalues f'. The Feynman propagator then is a matrix element of the unit operator $\langle f_2\sigma_2|1|f_1\sigma_1\rangle$. Wheeler[24] has introduced into his discussions of Feynman field theory the idea of representing the states explicitly as functionals $\psi_\sigma(f)$ of the *field configurations*. A field configuration is simply a specific value of the field given for every point of the surface σ. As an example of a state functional we quote from Wheeler the ground state of the electromagnetic field:

$$\psi_0 = N^{-1} \exp\left\{ -\frac{1}{16\pi^4\hbar c}\int\int \frac{\mathbf{H}(x)\cdot\mathbf{H}(y)}{|\mathbf{x}-\mathbf{y}|^2} d^3x d^3y \right\}. \tag{11a}$$

Here N is a normalization factor which is independent of the field, but which depends on the way the functional integration over field configurations is defined. This idea of a state functional is a key idea in the quantum theory of general relativity, for we may begin with state functionals and then later, when we know something about the inner product of two such functionals, find the relationship between these functionals and vectors in Hilbert space. In this way we avoid assuming that there exists a $g_{\mu\nu}(x)$ operator and a normalizable physical state which is an approximate eigenfunction of $g_{\mu\nu}(x)$. In a fully gauge-invariant theory neither of these assumptions is true.

The idea of a state functional is independent of any idea of an inner product. The idea of a state vector is that of an element of Hilbert space, and is therefore stated in terms of an inner product. To distinguish the two situations we use the notation ψ_a, ψ_b, for state functionals, and $|a\rangle$, $|b\rangle$ for the corresponding state vectors when they exist. A state vector $|a\rangle$ can be defined by a state functional ψ_a for which $\int|\psi_a(f)|^2\delta f$ is finite. This functional integration is over the field configurations f which are the arguments of ψ_a (not over field histories). It is easily possible that $|a\rangle = |b\rangle$ even though $\psi_a \neq \psi_b$. In (3) and (4) the quantities $(f_2\sigma_2|f_1\sigma_1)$ and $(f_2\sigma_2|\mathcal{O}|f_1\sigma_1)$ are functionals of the two field configurations f_2 and f_1. They serve to define matrix elements through the formula

$$\langle 2|\mathcal{O}|1\rangle = \int \psi_2^*(f_2)(f_2\sigma_2|\mathcal{O}|f_1\sigma_1)\psi_1(f_1)\delta f_2 \delta f_1, \tag{12}$$

where $|1\rangle$ and $|2\rangle$ are state vectors defined by the state functionals ψ_1 on σ_1 and ψ_2 on σ_2. We assign no meaning to the isolated symbol $|f\sigma\rangle$, and would wait until the existence of an appropriate state were made probable before writing $|f\sigma\rangle$.

Both Laurent and Wheeler employ a gauge-invariant action, $\int F_{\mu\nu}F^{\mu\nu}d^4x$, in quantizing the electromagnetic field, and suggest using $\int R(-g)^{\frac{1}{2}}d^4x$ in general relativity. This does not seem to avoid the use of a subsidiary condition, but does allow one to use a gauge invariant subsidiary condition. In electromagnetism this condition is

$$\psi_\sigma(A) = \psi_\sigma(A + \text{grad}\lambda) \text{ for all } \lambda. \tag{13}$$

In general relativity the subsidiary condition on the state functionals is similar in form: ψ_σ must be constant over every set of field configurations which differ only by gauge transformations.

3. NATURE OF QUANTUM THEORY OF GENERAL RELATIVITY

We now sketch an over-all view of what one is to mean by the quantization of general relativity with a minimum of explanation or justification:

(1) We introduce a *4-manifold*,[13] $\mathfrak{M}$, of points. These points have no physical significance in themselves. However, they serve as handles for stating the significant mathematical relationships. They act as formless building material; no metric has yet been introduced, nor will a definite choice of a metric for $\mathfrak{M}$ ever be made.

(2) We use subsets of these points called *hypersurfaces*, σ_1, σ_2, $\cdots$. They are 3-dimensional submanifolds[29] of $\mathfrak{M}$.

(3) We introduce the concept of a metric at a point x. The metric *is* the inner product operation $(\ ,\ )_x$. A particular metric is defined by giving a way of computing the inner product $u\cdot v \equiv (u,v)_x$ of every pair of tangent vectors u, v at x. It is most commonly defined by giving its components $g_{\mu\nu}(x)$ in a particular coordinate system so that the computation of $u\cdot v$ is

$$u\cdot v = (u,v) = g_{\mu\nu}u^\mu v^\nu.$$

[28] J. C. Polkinghorne, Proc. Roy. Soc. (London) A230, 272 (1955).

[29] C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, 1946). Definition 1, p. 85, and Def. 2, p. 80, but read differentiable for analytic.

The notation ds_x^2 for the metric at x is more common[30] than $(\ , \ )_x$ but has the same meaning. Thus we shall always mean by ds_x^2 a metric (an inner product operation) at x. In particular, the *operation* ds^2 may be used to compute distance intervals, but is *not* to be understood as itself being a distance interval.

(4) We introduce a *field history*, ds^2 for the metric, f for a more general field, by giving at each point of $\mathfrak{M}$ a definite value $ds^2(x)$ or $f(x)$ for the field there.

(5) If values $ds^2(x)$ or $f(x)$ are specified only over a single hypersurface σ_1, we call this a *field configuration* $(ds^2)_1$ or f_1 at σ_1.

(6) For any hypersurface σ or σ_1 we consider *state functionals* ψ_σ or ψ_1 (to be distinguished later from the narrower class of *physical state functionals*). A particular state functional ψ_1 at σ_1 is defined by giving for each field configuration f_1 at σ_1 a complex number $\psi_1(f_1)$, the value of ψ_1 for the field configuration f_1.

A sample state functional ψ_σ can be defined by

$$\psi_\sigma(ds^2) = \exp\left\{ -\int_\sigma {}^3R_{ij}\,{}^3R^{ij}({}^3g)^{\frac{1}{2}}d^3x \right\} \tag{14}$$

if σ is space-like for the metric ds^2,

$$\psi_\sigma(ds^2) = 0 \quad \text{otherwise.}$$

Here we assume coordinates around σ chosen so that σ is defined by $x^0(x)=0$. Having written the field configuration ds^2 on σ in the form

$$ds^2(x) = g_{00}(x)(dx^0)^2 + 2g_{0i}(x)dx^0dx^i + g_{ij}(x)dx^idx^j$$

we compute the three-dimensional Ricci curvature tensor ${}^3R_{ij}$ and volume element $({}^3g)^{\frac{1}{2}}d^3x$ from the metric ${}^3g_{ij} = g_{ij}$ defined on σ by ds^2. We do not know exactly what physical situation this sample state functional describes. It gives greatest probability to the field configurations ds^2 for which ${}^3R_{ij}=0$, i.e., for which the g_{ij} metric makes σ flat (assuming that the topology of σ allows a flat metric). It is assumed that ψ_1 gives a complete description of a physical state of the system, so that from ψ_1 *plus* the machinery of the theory one can compute the expectation value of every observable. In a particle theory $\mathfrak{M}$ is one dimensional, $-\infty < t < \infty$, and the σ are zero dimensional, $t=T$, while a state functional ψ_σ is an ordinary function $\psi_T(x)$ of the field $f = x =$ particle coordinate.

(7) From all conceivable *families* $\psi = \{\psi_t\}^{\infty}{}_{t=-\infty}$ of state functionals in particle theory, the Schroedinger equation, $i\hbar\partial\psi_t/\partial t = H\psi_t$, selects certain *admissible* ones called *states*. The entire admissible family of state functionals is determined when we give one member, say ψ_0. When we represent ψ by ψ_0 we are in the Heisenberg picture, while when we represent ψ by the admissible family $\{\psi_t\}$ we are in the Schroedinger picture.

A similar situation prevails in field theory where a state ψ may be represented either by a state functional ψ_0 at a chosen surface σ_0 (Heisenberg) or by a family $\{\psi_\sigma\}$ of state functionals (Schroedinger) satisfying the dynamical principle.

(8) As *dynamical principle* we use, in place of the Schroedinger equation, the Feynman principle: Given a state functional ψ_0 at σ_0, we find the other members of the family $\{\psi_\sigma\}$ for the state determined by ψ_0 through the formula

$$\psi_\sigma(f_\sigma) = \int K(f_\sigma,\sigma; f_0\sigma_0)\psi_0(f_0)\delta f_0. \tag{15}$$

Here $\int \cdots \delta f_0$ is a functional integration over field configurations f_0 at σ_0, and K is the Feynman propagator which may also be written $(f_\sigma\sigma | f_0\sigma_0)$.

(9) The Feynman propagator is to be defined by a functional integration over field histories:

$$K(f_2,\sigma_2; f_1,\sigma_1) = \int \exp\left\{ \frac{i}{\hbar} I_V(f) \right\} \delta f, \tag{16}$$

where $I_V(f) = \int_V \mathfrak{L} d^4x$ is the action integral for the classical theory of the field, evaluated for the particular field history f defined over the four-dimensional V which has σ_2 and σ_1 as boundaries. The functional integration extends over all field histories f with boundary values f_2 on σ_2 and f_1 on σ_1. We write

$$\delta f = N^{-1}(V) \prod_{x\in V} df_x \tag{17}$$

to suggest that the definition of the functional integration might be approached by the Wheeler-Feynman skeletonizing process (Sec. 2). Here df_x is a measure on the space of values $f(x) = f_x$ of the field at a point. The normalization factor $N(V)$ suggests the limiting processes and normalizations which go into the definition of the functional integration process.

In Sec. 4 we give two axioms (H) and (L) which guide in the choice of the measure df_x to be used in constructing δf, and in Sec. 6 we see that the operator field equations can be derived on the basis of these axioms. In Sec. 5 we define the measure $\int \cdots df_x$ to be used in quantizing general relativity, $f = ds^2$, and the operator field equations for general relativity are given explicitly in Sec. 6. The problem of defining the functional integral has been temporarily side-stepped, not solved.

(10) Gauge-invariance brings with it a subsidiary condition so that not every state functional ψ_0 at σ_0

[30] Computation methods more efficient than classical tensor analysis focus attention on the quadratic differential form $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ rather than upon its components $g_{\mu\nu}$ directly, [a] [b]. This viewpoint [c] leads to new insights into differential geometry. An intrinsic definition of a differential [b] [d] identifies the differential form ds^2 with the inner product operation $(\ , \ )$. [a] É. Cartan, *Leçons sur la Géométrie des Espaces de Riemann* (Gauthiers-Villars, Paris, 1946), second edition. [b] H. Flanders, Trans. Am. Math. Soc. **75**, 311 (1953). [c] S. S. Chern, Bull. Am. Math. Soc. **52**, 1 (1946). [d] See reference 29, p. 81.

generates a usable state $\psi=\{\psi_\sigma\}$ by (15). Those ψ_0 which do define usable states ψ, and the ψ they define, are called *physical states*. The criterion which distinguishes physical states is called the *subsidiary condition*. The sample state function of (14) satisfies the subsidiary condition in general relativity.

(11) Gauge invariance in general relativity has its origins in the *topological invariance* of the theory. Topological invariance means that, given two manifolds $\mathfrak{M}$ and $\mathfrak{M}'$ which are topologically equivalent (differentiably homeomorphic), if we first perform all constructions required by theory and then choose a particular homeomorphism (differentiable 1-1 correspondence between the points x of $\mathfrak{M}$ and x' of $\mathfrak{M}'$) between $\mathfrak{M}$ and $\mathfrak{M}'$, it will always be possible to extend it to a 1-1 correspondence between the elements of the structure over $\mathfrak{M}$ and those of the structure over $\mathfrak{M}'$. Roughly, topological invariance means that nothing in the theory changes on replacing $\mathfrak{M}$ by a topologically equivalent manifold. Thus we might let $\mathfrak{M}$ be the set of quadruples (r, z^1, z^2, z^3) where $r>0$ is a real number and each z^i is a complex number with $|z^i|=1$. On the other hand we might just as well let $\mathfrak{M}$ be the set of quadruples $\{s,\theta^1,\theta^2,\theta^3\}$ where s is any real number and each θ^i any family $\{\theta+2\pi n\}_n$ of real numbers $(n=\cdots -1, 0, 1, 2, \cdots)$. This second manifold is customarily called $R\times T^3$.

The topological invariance of general relativity has other consequences besides the subsidiary condition. If a pair of hypersurfaces σ_2, σ_1 (e.g., $s=2$, 1 in the foregoing) are topologically indistinguishable from another pair σ_4, σ_3 (e.g., $s=40$, 30), then the corresponding Feynman propagators $K(2,1)$ and $K(4,3)$ should be equivalent. This fact we consider in detail in Sec. 7 and find that it implies that the Hamiltonian operator vanishes. Consequently, among the members of a state $\psi=\{\psi_\sigma\}$ defined by a state functional ψ_0 on σ_0, every time there is no topologically invariant criterion which distinguishes the relationship of σ_1 to $\mathfrak{M}$ and σ_0 from the relationship of σ_2 to $\mathfrak{M}$ and σ_0, ψ_1 and ψ_2 will be equivalent. That is ψ_1 puts the same physics on σ_1 as ψ_2 puts on σ_2—ψ_σ is essentially independent of σ and the Heisenberg and Schroedinger pictures essentially coincide. For example, (14) defines a ψ_σ not just for a particular surface σ, but as σ is allowed to vary it defines a family $\psi=\{\psi_\sigma\}$. This family satisfies the dynamical principle (15) for the appropriate class of surfaces σ, e.g. those surfaces defined in $R\times T^3$ by $s=f_\sigma(\theta^1,\theta^2,\theta^3)$ where f_σ is any differentiable (periodic) function.

(12) The idea of Hamiltonian, Heisenberg, and Schroedinger picture, and many other concepts familiar in other field theories can only be incorporated in quantized general relativity by analogy. They do not have exactly the same meaning, for we do not have at hand any metric, or meaningful idea of time, to use in defining them. The manifold $\mathfrak{M}$ started out without a metric, and never gets one. We cannot, for example, speak of space-like surfaces, but must replace this idea with that of a *space-like state functional*: a ψ_σ which satisfies $\psi_\sigma(ds^2)=0$ for every field configuration ds^2 which does not make σ space-like.

The dynamical principle (15) was likewise written by analogy, but it turns out to be very undynamic (ψ_σ independent of σ)—at least we hope it would be undynamic in this sense if the constructions could be carried out. With no metric on $\mathfrak{M}$ there is little physical significance to a choice of σ so probability amplitudes ought not to depend on this choice. The true dynamics is the dependence of observables on time, and therefore takes place at a point in the theory where there is a metric with which a meaningful idea of time can be stated. The true dynamics is an *inner dynamics* which is to be found within the Feynman propagator, rather than in its external action on state functionals. We give one example: The volume $V(\sigma,ds^2)=\int_\sigma({}^3g)^{\frac{1}{2}}d^3x$ is an observable (gauge invariant) functional of σ and the field configuration ds^2 on σ. For a state ψ [such as the one defined by the state functional ψ_0 of Eq. (14) with σ_0 the surface $s=0$ in $R\times T^3$] its expectation value is

$$\langle V(\sigma)\rangle_\psi=\int\psi_\sigma^*(f)V(\sigma,f)\psi_\sigma(f)\delta f, \tag{18}$$

where the functional integration extends over all field configurations $f=ds^2$ at σ. Since the Hamiltonian vanishes, $\delta\langle V(\sigma)\rangle_\psi/\delta\sigma(x)=0$. This says not that $\langle V\rangle_\psi$ is a constant of motion, but that it is topologically invariant. To define the time derivative of V we need a field history, not just a field configuration. If ds^2 is given everywhere over V between σ_2 and σ_1 making σ_1 space-like, then we can define a function $t(x)$ as the length of that geodesic from x to σ_1 which is normal to σ_1. Then $t(x)$ is well defined for all x sufficiently near σ_1. Next define $\bar{V}(\sigma_1,ds^2,t)=V(\sigma_t,ds_t^2)$ where σ_T is the hypersurface $t(x)=T$, and ds_T^2 is the field configuration induced on σ_T by the field history ds^2. Then we have

$$\langle\dot{V}\rangle_\psi=\int\psi_2^*(f_2)\left[\frac{d}{dt}\bar{V}(\sigma_1,f,t)\right]_{t=0} \times\exp\left\{\frac{i}{\hbar}I_V(f)\right\}\psi_1(f_1)\delta f, \tag{19}$$

where ψ_2, ψ_1 are the members of $\psi=\{\psi_\sigma\}$ corresponding to σ_2 and σ_1, and the functional integration extends over all field histories $f=ds^2$ defined over V, including an integration over the boundary values f_n on σ_n. This is an admittedly complicated definition, for t depends on ds^2. We do not know how to evaluate such formulas, nor how to give them a mathematically precise meaning. This example is intended merely as an indication of how dynamical questions may be formulated within the theory.

4. H PRINCIPLE

We view the Feynman method as a Huygen's principle[31]: In forming

$$\int \exp\left\{\frac{i}{\hbar}\,(\text{action})\right\} d\,(\text{field histories}) \qquad (20)$$

the exponential by interference effects should select for us the fields that propagate significantly. In the integration all fields should be weighed equally, otherwise our arbitrary choice of what weight to put on each history in d (field histories) would compete with the exponential in selecting the significant fields. Of course we have yet to say what "weighed equally" means. The phrase suggests immediately, if we look about for suitable mathematical expression of it, the idea of invariant integration over a group.

Can we discover an invariant group integration in a simple problem? Consider a scalar field $\phi(x)$. An obvious group is the set of fields under addition. The integral over values of the field $\phi(x)=\phi_x$ at a point actually used here, $\int\cdots d\phi_x$, to generate the functional integration $\int\cdots\delta\phi$ is in fact the unique invariant integral over the additive group of real numbers ϕ_x. To construct $\int\cdots\delta\phi$ from $\int\cdots d\phi_x$, normalization factors $A(\mathfrak{N})$ appear depending on the grid of points $\mathfrak{N}$ used in the approximation of $\int\cdots\delta\phi$ by a multiple integral. The normalization factors are, however, independent of ϕ. In the Feynman propagator

$$K(\phi_2,\sigma_2;\phi_1,\sigma_1)=\int \exp\left\{\frac{i}{\hbar}I_V(\phi)\right\}\delta\phi, \qquad (16')$$

where V is the volume with σ_2 and σ_1 as boundaries and

$$I_V(\phi)=\int_V \mathfrak{L}d^4x$$

is the action integral, the functional integration $\int\cdots\delta\phi$ extends over all fields ϕ over V which reduce to ϕ_2 on σ_2 and ϕ_1 on σ_1. This set of fields, $F(V;\phi_2,\phi_1)$, does not form a group; however the fields ϕ over V which are zero on σ_2 and σ_1 do form a group, call it $G(V,bdyV)$, and the functional integration is invariant under this group. By this we mean

$$\int \mathcal{T}(\phi-\chi)\delta\phi=\int \mathcal{T}(\phi)\delta\phi \qquad (21)$$

for every functional $\mathcal{T}$ of ϕ [such as the exponential of the action in Eq. (16′)] when χ is in $G(V,bdyV)$. This invariance condition can be stated

$$\delta(\phi+\chi)=\delta\phi \quad \text{for} \quad \chi\epsilon G(V,bdyV)$$

as is seen by replacing ϕ by $\chi+\phi$ in the left-hand side of (21). The functional integral is thus invariant under a group $G(V,bdyV)$, even though it is an integral over $F(V;\phi_2,\phi_1)$. The relationship between G and F can be stated by saying that G is a transitive group of permutations[32] of F. That $\chi\epsilon G$ is a permutation of F means that the transformation χ of F defined by $\phi\rightarrow\phi+\chi$ is a 1-1 transformation of F onto F, while transitive means that if ϕ_0 is one field in F, every other $\phi\epsilon F$ is of the form $\phi=\phi_0+\chi$ for some $\chi\epsilon G$.

We now state these ideas in a still more abstract form in order to isolate certain aspects of the simple problem (quantizing a scalar field) and state them in a way which shows no traces of the problem in which they were recognized. Then the statements may be considered a discussion of the simpler aspects of more difficult problems. Were we to look immediately at general relativity the pertinent simplicities might get confused with difficult questions which we are not trying to answer immediately. Thus in the theory of an unspecified field f we will have a Feynman propagator

$$K(f_2,\sigma_2;f_1\sigma_1)=\int \exp\left\{\frac{i}{\hbar}I_V(f)\right\}\delta f \qquad (22)$$

defined by an action integral

$$I_V(f)=\int_V \mathfrak{L}d^4x \qquad (23)$$

and a functional integration $\int\cdots\delta f$. The integral extends over the set $F(V;f_2,f_1)$ of fields which reduce to f_2 on σ_2 and f_1 on σ_1. We want to have a transitive group of permutations of $F(V;f_2,f_1)$; we denote this group by $G(V,bdyV)$.

Definition: If G is a transitive group of permutations of F we say

F is an *homogeneous* set and G its *homogeneity*.

If $\pi\epsilon G$ is one of the transformations, the transform of $f\epsilon F$ will be written f^π in order not to suggest that we are necessarily dealing with a linear theory. (In the preceding example $f=\phi$, $\pi=\chi$, and $f^\pi=\phi+\chi$.) Successively applying two transformations, π then ω, we get $(f^\pi)^\omega=f^{\pi\omega}$ where $\pi\omega$ is the product in G. (In the example the group was written additively and $\pi\omega=\chi_1+\chi_2$ if $\pi=\chi_1$, $\omega=\chi_2$.) The invariance of the functional integration under G is expressed by

$$\delta(f^\pi)=\delta f \qquad (H)$$

which states that the volume element at any point f^π of F is just as big as it is at f, and means that for any

[31] R. P. Feynman, reference 2, Sec. 7, p. 377.

[32] This is identical with the idea of a transitive transformation group, but this name is more common when topological considerations also enter. D. Montgomery and L. Zippin, *Topological Transformation Groups* (Interscience Publishers, Inc., New York, 1955), Sec. 1.19, 1.26, 2.12.

functional $\mathcal{T}$

$$\int \mathcal{T}(f^{\pi^{-1}})\delta f = \int \mathcal{T}(f)\delta f \tag{24}$$

for all $\pi \epsilon G$.

This *H principle* is so named because of the Huygen's principle which suggested the "weighed equally" phase, the Homogeneity[32,33] of the fields F under the group G which gives precise meaning to the phrase, and the Hurwitz integral[34] or Haar measure[33,35] which will enter into the construction of an integration satisfying this postulate.

As a matter of convenience, we note that an invariant integral

$$\mathcal{J}(\mathcal{T}) = \int \mathcal{T}(f)\delta f \tag{25}$$

over $F(V; f_2, f_1)$ can be defined (formally) in terms of an invariant integral $\int \cdots \delta\pi$ over $G(V, bdyV)$. We define $\mathcal{T}^\pi$, the transform of a functional $\mathcal{T}$ under π, by the condition that $\mathcal{T}^\pi$ have the same value at f^π as $\mathcal{T}$ did at f, i.e.

$$\mathcal{T}^\pi(f^\pi) = \mathcal{T}(f). \tag{26}$$

Then the invariance requirement (24) can be stated

$$\mathcal{J}(\mathcal{T}^\pi) = \mathcal{J}(\mathcal{T}) \quad \text{for all } \pi \epsilon G(V, bdyV). \tag{27}$$

We define $\mathcal{J}$ by

$$\mathcal{J}(\mathcal{T}) = \int \mathcal{T}(f_0^\pi)\delta\pi, \tag{28}$$

where f_0 is some fixed field.
Then

$$\mathcal{J}(\mathcal{T}^\omega) = \int \mathcal{T}^\omega(f_0^\pi)\delta\pi = \int \mathcal{T}(f_0^{\pi/\omega})\delta\pi$$

$$= \int \mathcal{T}(f_0^\pi)\delta(\pi\omega) = \mathcal{J}(\mathcal{T})$$

provided $\delta(\pi\omega) = \delta\pi$, i.e., if $\int \cdots \delta\pi$ is a right invariant integral over $G(V, bdyV)$. This definition of $\mathcal{J} = \int \cdots \delta f$ in terms of $\int \cdots \delta\pi$ appears to depend on the choice of f_0; however if $\int \cdots \delta\pi$ is left invariant, $\delta(\omega\pi) = \delta\pi$, then

$$\mathcal{J}(\mathcal{T}) = \int \mathcal{T}(f_0^\pi)\delta\pi = \int \mathcal{T}(f_0^{\omega\pi})\delta\pi.$$

Thus f_0^ω, i.e., any field in $F(V; f_2, f_1)$, could have been used in place of f_0 without changing the value of $\mathcal{J}(\mathcal{T})$.

In view of the construction of $\int \cdots \delta f$ we imagine (17), the homogeneity group of the functional integral is to be generated by an homogeneity group $G(x)$ of the point integral $\int \cdots df_x$ over the space $F(x)$ of field values $f_x = f(x)$ at x. If $\pi(x) \epsilon G(x)$ we write the image of $f(x)$ under $\pi(x)$ as $f(x)^{\pi(x)}$. Let $G(x)$ be an homogeneity of $F(x)$ for each x, and π a field with values $\pi(x) \epsilon G(x)$ and for x on $bdyV$, $\pi(x) = 1$. Then we define the homogeneity $G(V, bdyV)$ of the set of fields $F(V; f_2, f_1)$ as the group of all transformations $f \rightarrow f^\pi$ where f^π is defined by

$$f^\pi(x) = f(x)^{\pi(x)} \quad \text{for each } x \epsilon V. \tag{L}$$

If π, ω are two elements of $G(V, bdyV)$ their product is defined by

$$(\pi\omega)(x) = \pi(x)\omega(x)$$

where the product on the right is taken in $G(x)$ for each $x \epsilon V$. Eq. (L) says that the π transformations are Local. Thus we have reduced the problem defining $G(V, bdyV)$ to that of defining $G(x)$. Since $\int \cdots \delta f$ can be defined formally in terms of $\int \cdots \delta\pi$, a point integral $\int \cdots d\pi_x$ over $G(x)$ can be taken as the starting point for a construction of the Feynman propagator.

5. MEASURE ON METRICS

Applying the ideas of the preceding section, we must define a transitive group of permutations $G(x)$ on the space $F(x)$ of metrics at a point. If $g_{\mu\nu}(x)$ are the components of $ds^2(x)$ we define the components $g_{\mu\nu}{}^\pi(x)$ of $ds^2(x)^{\pi(x)}$ by

$$g_{\mu\nu}{}^\pi(x) = \pi_\mu{}^\alpha(x) g_{\alpha\beta}(x) \pi_\nu{}^\beta(x), \tag{29}$$

where $\pi(x)$ is a tensor satisfying

$$\det\pi = \frac{1}{4!}\delta_{\alpha\beta\gamma\delta}{}^{\mu\nu\rho\sigma}\pi_\mu{}^\alpha\pi_\nu{}^\beta\pi_\rho{}^\gamma\pi_\sigma{}^\delta > 0. \tag{30}$$

The point integral[36]

$$\int \cdots d\pi_x = \int \cdots (\det\pi)^{-4} \prod_{\mu,\nu=0}^{3} d\pi_\nu{}^\mu \tag{31}$$

is a left and right invariant integral over $G(x)$, unique to within a factor independent of π. We *assume* that a functional integral $\int \cdots \delta\pi$ will be generated by $\int \cdots d\pi_x$. Since in $\delta\pi = N^{-1}(V)\prod d\pi_x$ the normalization factor is independent of π, $\int \cdots \delta\pi$ will be a left and right invariant integral over $G(V, bdyV)$, and thus give an invariant functional integral over $F(V; f_2, f_1)$ by (28). This $\int \cdots \delta f$ is then used in (22) to define the Feynman propagator of general relativity. The action integral we use in general relativity is

$$I_V(ds^2) = \frac{c^3}{16\pi G}\int_V R(-g)^{\frac{1}{2}} d^4x. \tag{32}$$

---

[33] A. Weil, *L'Intégration dans les Groupes Topologiques* (Hermann et Cie, Paris, 1940), Chap. II.

[34] E. P. Wigner, *Gruppentheorie* (Edwards Brothers, Inc., Ann Arbor, 1944), pp. 103–108.

[35] P. R. Halmos, *Measure Theory* (D. Van Nostrand Company, Inc., New York, 1950), pp. 254, 263.

[36] F. D. Murnaghan, *The Theory of Group Representations* (Johns Hopkins Press, Baltimore, 1938), p. 204.

To what extent is the choice of $G(x)$ arbitrary? We do not know precisely,[37] however since we have gone to all this trouble to avoid problems raised by gauge-invariance, it is worth noting that $G(x)$ coincides with the gauge group at the single point x (consider $\pi_\nu{}^\mu = \partial y^\mu/\partial x^\nu$). This criterion on the choice of $G(x)$—let $G(x)$ be the gauge group at x—also gives the right answer in electromagnetic theory and determines $G(x)$ uniquely. The choice of $G(x)$ should be independent of the Lagrangian of the theory however. It is much the same question as "Is the field f to be a scalar, vector, tensor, or what?" which must be decided in a classical theory before field equations are written for the theory. If we erroneously thought of general relativity as a theory of simply a symmetric tensor field $g_{\mu\nu}$, then the space $\bar{F}(x)$ of field values $g_{\mu\nu}(x)$ at x consists of all symmetric tensors and is a vector space under addition. The obvious choice of an homogeneity $\bar{G}(x)$ of $(x\bar{F})$ is the group of translations $g_{\mu\nu}{}^\pi = g_{\mu\nu} + \pi_{\mu\nu}$ and linear integration

$$\int \cdots \prod_{\mu\le\nu} dg_{\mu\nu} \qquad (33)$$

is unique to within a factor independent of $g_{\mu\nu}$. Using $\int R(-g)^{\frac{1}{2}}d^4x$ as the action integral in the Feynman propagator, and $\bar{G}$ as the homogeneity we could probably construct a quantum theory with the Einstein equations

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = 0 \qquad (34)$$

as its classical limit. However, neither the action integral, nor the group $\bar{G}$, distinguishes between the (diagonal) tensors $\hat{g}_{\mu\nu} = [-1,1,1,1]$ and $g_{\mu\nu}{}^* = [1,1,1,1]$. Both are solutions of the (34), and neither would have a preferred status in the classical limit of the quantum theory based on $\bar{G}$, so this classical theory would not be general relativity.

The invariant integral (31) over $G(x)$ is unique to within a factor independent of π. We have chosen this factor in (31) in a definite way that is independent of the choice of coordinates used to represent π by its components $\pi^\mu{}_\nu$. The invariant integral (33) over $\bar{F}(x)$ is also unique to within a factor independent of $g_{\mu\nu}$. In this case, however, had we chosen different coordinates $\bar{x}^\mu(x^\nu)$ we would have written

$$\prod_{\mu\le\nu} d\bar{g}_{\mu\nu} = r(A) \prod_{\mu\le\nu} dg_{\mu\nu}, \qquad (35)$$

where $A = \|\partial x^\mu/\partial\bar{x}^\nu\|$ and r is some function of the matrix A. It is easily seen that r defined in this way is a representation of the general linear group. As a one-dimensional representation it is necessarily of the form[38]

$$r(A) = (\det A)^k$$

and by counting the number of factors $A_\nu{}^\mu$ which enter in (35) from the transformation law for $g_{\mu\nu}$ we see that $k=5$. This observation allows us to write down immediately the integral

$$\int \cdots d(\text{metric})_x = \int \cdots (-g)^{-\frac{5}{2}} \prod_{\mu\le\nu}{}^{\frac{1}{2}} dg_{\mu\nu} \qquad (36)$$

over $F(x)$ invariant under $G(x)$, where again the choice of the arbitrary $g_{\mu\nu}$-independent factor is made independently of the choice of coordinates. The range of integration in (36) is the connected region of $g_{\mu\nu}$-space containing $g_{\mu\nu} = \text{diag}\ (-1, 1, 1, 1)$ and bounded by $\det g_{\mu\nu} = 0$.

6. OPERATOR FIELD EQUATIONS

Consider the integral (22) for the propagator in the form

$$K(f_2,\sigma_2; f_1,\sigma_1) = \int \exp\{iI_V(f^\pi)/\hbar\}\delta f. \qquad (37)$$

Using the invariance (H) of the integration we have inserted a transformation $\pi\epsilon G(V,bdyV)$. This integral is a functional of π which is independent of π, therefore

$$0 = \frac{\delta K}{\delta\pi(x)} = \frac{i}{\hbar}\int \frac{\delta I_V(f^\pi)}{\delta\pi(x)} \exp\left\{\frac{i}{\hbar} I_V(f^\pi)\right\}\delta f. \qquad (38)$$

Here the requirement (L) that the π's be local transformations of the field f was used, since π must be a field $\pi(x)$ in order to define the variational derivatives in (38). Defining the Jacobian transformation $\partial f^\pi/\partial\pi$ by

$$\delta f^\pi(x) = \frac{\partial f_x{}^\pi}{\partial \pi_x}\delta\pi(x), \qquad (39)$$

Eq. (38) can be written

$$\left\langle 2 \left| \frac{\delta I_V(f)}{\delta f(x)} \left(\frac{\partial f_x{}^\pi}{\partial\pi_x}\right)_{\pi=1} \right| 1 \right\rangle = 0 \qquad \text{(QF)}$$

after setting π = the identity transformation. These are the Quantum mechanical form of the Field equations. The equations

$$\frac{\delta I_V(f)}{\delta\pi(x)} = \frac{\delta I_V(f)}{\delta f(x)}\frac{\partial f_x{}^\pi}{\partial \pi_x} = 0 \qquad (40)$$

---

[37] The problem then is: Are there two distinct Lie groups G_1, G_2, which are effective homogeneities of the coset space $F = Gl(4)/L$ considered simply as a differentiable manifold, and which define essentially different invariant integrals on F. (L = Lorentz group.)

[38] We thank Professor V. Bargmann for pointing out this argument.

are equivalent to the Classical Field equations

$$\frac{\delta I_V(f)}{\delta f(x)}=0 \qquad \text{(CF)}$$

for the Jacobian is of maximum rank: any desired change in f^π may be produced by a suitable change in π. (G is transitive.)

In linear field theories equations (40) are identical to (CF); it follows directly from $f^\pi=f+\pi$ (linearity) that $\partial f^\pi/\partial\pi=1$.

In general relativity we have from (29) a representation of the Jacobian transformation by a matrix

$$\frac{\partial g^\pi{}_{\mu\nu}}{\partial\pi_\beta{}^\alpha}=g_{\alpha\gamma}\pi_\nu{}^\gamma\delta_\mu{}^\beta+\pi_\mu{}^\nu g_{\gamma\alpha}\delta_\nu{}^\beta \qquad (41)$$

when π is the identity, $\pi_\beta{}^\alpha=\delta_\beta{}^\alpha$ and

$$\left(\frac{\partial g_{\mu\nu}{}^\pi}{\partial\pi_\beta{}^\alpha}\right)_{\pi=1}=g_{\alpha\nu}\delta_\mu{}^\beta+g^\mu{}_\alpha\delta_\nu{}^\beta. \qquad (42)$$

The field equations (CF) in this coordinate system are

$$\frac{\delta I(ds^2)}{\delta g_{\mu\nu}(x)}=\frac{-c^3}{16\pi G}(R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R)(-g)^{\frac{1}{2}} \qquad (43)$$

so (QF) reads

$$\langle 2|(R_\alpha{}^\beta-\tfrac{1}{2}\delta_\alpha{}^\beta R)(-g)^{\frac{1}{2}}|1\rangle=0. \qquad (44)$$

Obvious linear combinations of these equations give

$$\langle 2|R(-g)^{\frac{1}{2}}|1\rangle=0 \qquad (45)$$

and

$$\langle 2|R_\alpha{}^\beta(-g)^{\frac{1}{2}}|1\rangle=0. \qquad (46)$$

The form of (44) depends only upon the invariance of the integration with respect to the transformation π, not upon the fact that we concern ourselves with the $g_{\mu\nu}$ rather than $g^{\mu\nu}$ or some other equivalent tensor constructable algebraically from $g_{\mu\nu}$. A π transformation of ds^2 defines corresponding transformations of these tensors as well, and $\delta I_V/\delta\pi(x)$ is independent of the factorization (40) used to compute it.

The content of (44) is this: in the quantum theory of general relativity there exists an operator corresponding to the classical tensor $(R_\alpha{}^\beta-\frac{1}{2}\delta_\alpha{}^\beta R)(-g)^{\frac{1}{2}}$, and this operator is the zero operator. There is in (44) *no* statement about operators corresponding to $(-g)^{\frac{1}{2}}$ nor $R_\alpha{}^\beta-\frac{1}{2}\delta_\alpha{}^\beta R$; hence the order in which the factors are written in Eq. (44) is not significant.

7. TOPOLOGICAL INVARIANCE

Conventional field theories start with a manifold possessing a metric (usually the flat Minkowski metric) and describe the way in which some other field is to be assigned to the manifold. Such theories are metrically invariant (Lorentz invariant): they are unaffected by any transformation of the manifold which leaves the metric unchanged. General relativity starts with a manifold and describes the way in which a metric is to be assigned to the manifold. In general relativity, only the differentiable structure (frequently abbreviated "topology") is fixed at the start, and the theory of general relativity is unaffected by any transformation of the manifold which leaves the topology unchanged. We do not mean that every quantity in the theory remains absolutely fixed, but that to every permissible transformation of the manifold there corresponds a transformation of all the other quantities in the theory of such a nature as to leave the physical content unchanged. A transformation of a manifold which leaves the topology unchanged is called a homeomorphism. In this section we look at the transformation of the Feynman propagator corresponding to a homeomorphism, and see that if every homeomorphism is a permissible transformation (i.e., if the theory is topologically invariant) as in the theory of general relativity, then the Feynman propagator connecting equivalent hypersurfaces is trivial.

We say two hypersurfaces σ_2 and σ_3 are *equivalent* if (1) there exists a hypersurface σ_1, a V_{21} with σ_2 and σ_1 as boundaries, and a V_{31} with σ_3 and σ_1 as boundaries, and (2) there is a homeomorphism h of the 4-manifold $\mathfrak{M}$ which maps V_{31} onto V_{21}, σ_3 onto σ_2, and which restricts to the identity map of σ_1 onto itself.

A homeomorphism gives not only a correspondence between points by $y=\mathrm{h}x$, but can be used to construct a correspondence of all fields. If ϕ is a scalar function, the corresponding function under h is written ϕ^{h} and defined by $\phi^{\mathrm{h}}(x)=\phi(\mathrm{h}x)$. The metric corresponding under h to ds^2 will be designated by $(ds^2)^{\mathrm{h}}$. If x^μ, y^μ are local coordinates around x, and around $y=\mathrm{h}x$, respectively, then h can be represented locally by differentiable functions h^μ:

$$y^\mu=\mathrm{h}^\mu(x^0,x^1,x^2,x^3). \qquad (47)$$

The components $g_{\mu\nu}{}^{\mathrm{h}}(x)$ of $(ds^2)^{\mathrm{h}}(x)$ are given in terms of the components $g_{\mu\nu}$ of ds^2 by

$$g_{\mu\nu}{}^{\mathrm{h}}(x)=g_{\alpha\beta}(\mathrm{h}x)\frac{\partial\mathrm{h}^\alpha}{\partial x^\mu}\frac{\partial\mathrm{h}^\beta}{\partial x^\nu}. \qquad (48)$$

In a similar way a field f^{h} corresponding under h to a field f could be defined. A state functional $\mathrm{h}\psi$ at $\mathrm{h}\sigma$ corresponding to ψ at σ is then defined by requiring

$$\mathrm{h}\psi(f)=\psi(f^{\mathrm{h}}) \qquad (49)$$

for every field configuration f at $\mathrm{h}\sigma$.

Let h be a homeomorphism which establishes the equivalence of two hypersurfaces σ_2 and σ_3. A state functional ψ_1 at σ_1 (see definition of equivalent) defines a state (family of state functionals, see Sec. 3)

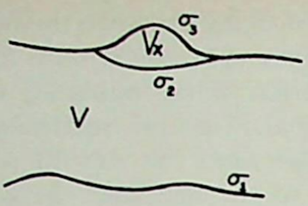


FIG. 1. The vanishing of the Hamiltonian. The hypersurface σ_3 is obtained from σ_2 by a change $\delta\sigma(x)$ which consists of sweeping σ_2 across V_x into σ_3. The corresponding change $\delta\psi_2$ in the state functional defines the Hamiltonian density operator $\mathcal{H}_{\text{op}}(x)$ according to the Schwinger-Tomonaga equation

$$\delta\psi_2(\bar{f}_2)=(i/\hbar)\int(\bar{f}_2|\int_{V_x}\mathcal{H}(x)d^4x|f_2)\psi_2(f_2)\delta f_2.$$

The state is defined by specifying a state functional ψ_1 on σ_1. Standing at σ_1 however, we are unable (in a topologically invariant theory) to formulate a criterion for choosing a second hypersurface σ which would permit the choice $\sigma=\sigma_2$ but which would not allow $\sigma=\sigma_3$. Thus ψ_2 and ψ_3 are equivalent, i.e. $\delta\psi_2=0$. In any topologically invariant theory we therefore must have $\mathcal{H}_{\text{op}}=0$.

containing the state functionals ψ_2 at σ_2 and ψ_3 at σ_3 which are defined by ψ_1 and the Feynman propagator over V_{21} and V_{31}, respectively. In a "matrix" notation which suppresses arguments (f_n) and integrations (δf_1) we may write (compare (15))

$$\psi_2=K(\sigma_2,\sigma_1)\psi_1$$
$$\psi_3=K(\sigma_3,\sigma_1)\psi_1. \qquad (50)$$

Since h is the identity on σ_1, $\text{h}\sigma_1=\sigma_1$ and $\text{h}\psi_1=\psi_1$. If h is a permissible transformation we must have

$$\text{h}K(\sigma_3,\sigma_1)=K(\text{h}\sigma_3,\text{h}\sigma_1)\text{h}, \qquad (51)$$

or in the present case, since $\text{h}\sigma_3=\sigma_2$,

$$\text{h}\psi_3=\psi_2. \qquad (52)$$

In summary, let $\psi=\{\psi_\beta\}$ be a state for a topologically invariant quantum theory, and let σ_2, σ_1 be equivalent under a homeomorphism h with $\text{h}\sigma_2=\sigma_1$, then $\text{h}\psi_2=\psi_1$. The equation $\psi_2=\psi_1$ is nonsense since ψ_2 and ψ_1 are functionals with different argument domains (the field configurations on σ_2 and σ_1, respectively). In a topologically invariant theory the only distinguished transformations which we could use to identify these two domains are the homeomorphisms. Consequently one does not misinterpret the equation $\text{h}\psi_2=\psi_1$ by saying that ψ_2 is the same as ψ_1. When ψ_2 is given by an explicit formula such as (14), the equation $\text{h}\psi_2=\psi_1$ says that a formula for ψ_1 may be obtained simply by replacing σ_2 by σ_1 in the formula for ψ_2. The general idea is that in a topologically invariant theory, a homeomorphism is the same as an identity transformation, modulo physically meaningless mathematical distinctions.

We have taken as the basic elements of quantum field theory the definition (16) of the Feynman propagator, the dynamical principle (15), and the formula (4) for computing matrix elements. We did not have any need for a Hamiltonian in the classical theory (in general relativity it is quite difficult to find one[7]) so we may define a Hamiltonian density operator $\mathcal{H}_{\text{op}}(x)$ directly in the quantum theory. The Schwinger-Tomonaga equation,[39,40]

$$i\hbar\frac{\delta\psi_\sigma}{\delta\sigma(x)}=\mathcal{H}_{\text{op}}(x)\psi_\sigma, \qquad (53)$$

serves this purpose. Let a pair of hypersurfaces such as σ_2 and σ_3 in Fig. 1 differ by the boundary of a small (cellular) neighborhood V_x of a point x. A comparison of ψ_3 and ψ_2 is defined by

$$\delta\psi_2=\text{h}\psi_3-\psi_2, \qquad (54)$$

where h is any homeomorphism $\text{h}:\sigma_3\rightarrow\sigma_2$ which reduces to the identity on the common submanifolds of σ_3 and σ_2. There is no structure available in a topologically invariant theory to make a more definite choice of h, but $\mathcal{H}_{\text{op}}$ will be well defined without any further restrictions on h. Note that the definition of $\text{h}\psi_3$ in terms of ψ_3 is independent of the equations of motion (15). Now suppose that both ψ_3 and ψ_2 are members of a state ψ defined by a state functional ψ_1 on σ_1, and that V is bounded by σ_2 and σ_1 so $V+V_x$ is bounded by σ_3 and σ_1. We may in this case interpret $\delta\psi_2$ as the change in the state functional produced by a change $\delta\sigma(x)$ displacing the surface σ_2 across V_x. Then since V_x is assumed to be a cell (a homeomorph of the unit cube E^4 in Euclidean 4-space) the homeomorphism $\text{h}:\sigma_3\rightarrow\sigma_2$ can be extended to a homeomorphism $\text{h}:V+V_x\rightarrow V$ which is the identity on σ_1. Therefore σ_3 and σ_2 are equivalent, $\text{h}\psi_3=\psi_2$, $\delta\psi_2=0$, and consequently $\mathcal{H}_{\text{op}}=0$. *In any topologically invariant theory, the Hamiltonian operator vanishes.* For the conclusion $\mathcal{H}_{\text{op}}=0$ to hold it is necessary that the quantum theory, not just the classical theory, be topologically invariant.

Because of different methods of definition, our $\mathcal{H}_{\text{op}}$ Hamiltonian is not necessarily the operator corresponding to the Hamiltonian defined in classical theory. However Professor J. L. Anderson at the Chapel Hill conference voiced suspicions that the classical Hamiltonian in general relativity would be zero.

Is the vanishing of the Hamiltonian embarrassing? Not at all. As defined here the Hamiltonian is, so to speak, the generator of infinitesimal homeomorphisms and is totally unrelated to the energy of the system or to the dynamic changes taking place in the observables. Its vanishing is merely the local statement of the topological invariance of the theory, and is therefore desirable. We postulate it as one of the characteristic properties of the Feynman propagator in quantizing general relativity or any other topologically invariant field theory.

SUMMARY OF CONCRETE RESULTS

It may reasonably be expected that whenever any construction of the Feynman integral is possible,

[39] J. Schwinger, Phys. Rev. 74, 1449 (1948).
[40] S. Tomonaga, Progr. Theort. Phys. 1, 34 (1946).

this construction can be done in conformity with the H principle. The H principle characterizes a property of the Feynman integral sufficient to guarantee that an operator form of the field equations will hold. The choice of the homogeneity of the field determines uniquely the relationship between the form of the operator field equations and the form of the classical field equations. The homogeneity of the metric field of general relativity has been defined, and the operator form of the Einstein field equations has been given. In a topologically invariant quantum field theory, such as quantized general relativity, the (Schroedinger) state functionals on equivalent hypersurfaces are equivalent, so that the Hamiltonian vanishes.

ACKNOWLEDGMENTS

Special appreciation is expressed to the sponsors of the International Conference on the Role of Gravitation in Physics held at Chapel Hill, North Carolina, January 18–22, 1957, and to Professor and Mrs. Bryce DeWitt for the stimulus which this conference provided. Among the many participants whose remarks helped to clarify the Feynman quantum theory of general relativity, we are particularly indebted to Professor J. L. Anderson, Dr. S. Deser, and Dr. B. Laurent. Above all, we thank Professor J. A. Wheeler. He originally suggested the problem of formulating the Feynman quantization of general relativity. We regret that our approach to this problem has not the conciseness of his solution to it, which was simply to write

$$\int \exp\{(i/\hbar)(\text{Einstein action})\}d\ (\text{field histories}).$$

He provided constant encouragement in pursuing the investigation; and his provoking ideas on the nature of the theory of general relativity led us to adopt the intrinsic or coordinate-free point of view which has been essential in the development of this theory. His help and advice in writing this paper are reflected in those sections which are most clearly presented.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

Reality of the Cylindrical Gravitational Waves of Einstein and Rosen

JOSEPH WEBER, *Lorentz Institute, University of Leiden, Leiden, Netherlands, and University of Maryland, College Park, Maryland*

AND

JOHN A. WHEELER, *Lorentz Institute, University of Leiden, Leiden, Netherlands, and Palmer Physical Laboratory, Princeton University, Princeton, New Jersey*

1. INTRODUCTION

QUESTIONS have been raised whether gravitational radiation has any well-defined existence.[1] Supporting this skeptical position, Rosen has investigated further[2] the cylindrical gravitational waves first considered by him and Einstein[3] as an outgrowth of a suggestion by H. P. Robertson. A monochromatic wave, or a pulse, of cylindrical symmetry, moves inward in matter-free space, implodes on the axis, and moves out again. This is the only problem of gravitational radiation where one has an *accurate* solution of the field equations of general relativity. The problem is special enough not to illustrate all features of gravitational radiation. On the other hand, all correct general statements about gravitational radiation must obviously be compatible with this problem. This problem therefore occupies a special position in the theory of gravitational radiation.

Rosen finds an unexpected result. The pseudotensor that measures the density of gravitational energy and momentum in the cylindrical wave is everywhere zero. The significance of this finding is the subject of this paper. We conclude that many of the otherwise apparently paradoxical properties of this cylindrical wave can be understood by taking into account the analogy between gravitational waves and electromagnetic waves, and the special demands of the equivalence principle, which rules out a special role for any particular frame of reference.

Section 2 recapitulates the expressions of Einstein and Rosen and of Rosen for the metric of the cylindrical wave. Two kinds of solution are of interest: monochromatic waves and pulses. A pulse type of solution is constructed that is represented by particularly simple mathematical expressions. Section 3 reviews the proof that the pseudotensor density of gravitational

[1] A. E. Scheidigger, Revs. Modern Phys. 25, 451 (1953).

[2] N. Rosen in *Jubilee of Relativity Theory*, edited by A. Mercier and M. Kervaire (Birkhäuser Verlag, Basel, 1956).

[3] A. Einstein and N. Rosen, J. Franklin Inst. 223, 43 (1937); N. Rosen, Bull. Research Council Israel 3, 328 (1953).

momentum and energy is everywhere zero, both for monochromatic waves and for pulses. From this it does not follow that the energy of the gravitational wave is zero. Rather, this energy is *undefined* because the wave is unbounded. Were the pulse imploding, not on an infinite straight line, but on a circular line, the energy would be expected to be finite and reasonable, according to simple physical arguments. Section 4 returns to the infinitely extended case where the exact solutions are available. The curvature tensor of Riemann, R_{ijkl}, does not vanish in the region occupied by the wave; nor do the invariants formed from the components of this curvature tensor vanish. The disturbance in question is real and not removable by any change of coordinate system. In Sec. 5 influence of the gravitational wave upon the motion of an infinitesimal test particle, at rest before the arrival of a pulse, is studied. The particle receives a velocity from the ingoing pulse, but after implosion the pulse returns with reversed force and reduces the velocity to zero. The analogies are stressed with the case of a charged particle subject to the action of an electromagnetic plane wave. There the passing wave gives no net forward impulse to the particle in the approximation in which the secondary wave radiated by the particle itself is neglected. When allowance is made for radiation reaction, the equation of motion of the particle predicts a forward push. Several predictions about the response of a test particle to gravitational radiation are made which are in principle subject to check. It is concluded that the cylindrical waves of Einstein and Rosen present no real paradox; that the apparently anomalous behavior of these waves is completely consistent with the equivalence principle of general relativity.

2. CYLINDRICAL WAVES

Spherical gravitational waves, like spherical electromagnetic waves, can never be truly spherically symmetric. The polarization tensor in the one case, or the polarization vector in the other case, cannot keep a constant magnitude and change continuously in direction over the surface of a sphere, according to the fixed point theorem of topology. Cylindrical gravitation waves, like cylindrical electromagnetic waves, avoid this difficulty. The polarization at a point is described by a small ellipse normal to the direction of wave propagation. This ellipse describes the distances from a central test particle of a set of test particles which were originally at rest upon a circle before the wave fell upon them. The principal axes of this ellipse are parallel and perpendicular to the axis of the cylinder. As the phase of the oscillation advances, the long axis becomes the short one, and conversely, but the axes do not rotate. The cylindrical waves of Einstein and Rosen make no use of the other independent state of polarization, in which the principal axes are turned at 45° to the axis of the cylinder.

A typical point is designated by cylindrical coordinates, ρ, φ, and z. The product of the velocity of light and the time is represented by the cotime, T. In these coordinates the metric of Einstein and Rosen has the form

$$ds^2=e^{2\gamma-2\psi}(-dT^2+d\rho^2)+\rho^2e^{-2\psi}d\varphi^2+e^{2\psi}dz^2 \qquad (1)$$

where the dilatation quantities, γ and ψ, are functions of ρ and T alone. The gravitational field equations are satisfied in empty space for a metric of the form (1) provided ψ and γ satisfy the equations,

$$\psi_{\rho\rho}+\frac{1}{\rho}\psi_\rho-\psi_{TT}=0, \qquad (2)$$

$$\gamma_\rho=\rho[\psi_\rho^2+\psi_T^2], \qquad (3)$$

$$\gamma_T=2\rho\psi_\rho\psi_T. \qquad (4)$$

Solutions of (2) are well-known and represent cylindrical waves. It was originally thought that these waves are capable of transmitting energy to unlimited distances because one can write a solution representing progressive waves, where the first dilatation factor has the form

$$\psi=AJ_0(\omega\rho)\cos\omega t+BN_0(\omega\rho)\sin\omega t; \qquad (5)$$

and the second factor, in the special case $B=A$, reduces to the expression

$$\begin{aligned}\gamma=\tfrac{1}{2}A^2\omega\rho\{&J_0(\omega\rho)J_0'(\omega\rho)+N_0(\omega\rho)N_0'(\omega\rho)\\&+\omega\rho[(J_0(\omega\rho))^2+(J_0'(\omega\rho))^2+(N_0(\omega\rho))^2\\&+(N_0'(\omega\rho))^2]+[J_0(\omega\rho)J_0'(\omega\rho)-N_0(\omega\rho)N_0'(\omega\rho)]\\&\times\cos 2\omega T+[J_0(\omega\rho)N_0'(\omega\rho)+N_0(\omega\rho)J_0'(\omega\rho)]\\&\times\sin 2\omega T\}-\frac{2}{\pi}A^2\omega T. \qquad (6)\end{aligned}$$

The last term of (6) is aperiodic in time. It leads to a systematic change of the metric with time which was at first interpreted as due to loss of energy. Rosen gave arguments against this interpretation.[3] With Rosen, we exclude solutions that contain the irregular Bessel function, $N_0(\omega\rho)$, as not well defined at the origin.

More interesting than a monochromatic dilatation factor ψ of the form (5), with $B=0$, is a pulse formed by linear superposition of such waves.[3] We superpose such waves with an amplitude factor $A=2Ce^{-a\omega}$, thus:

$$\begin{aligned}\psi&=2C\int_0^\infty e^{-a\omega}J_0(\omega\rho)\cos\omega T d\omega\\&=C[(a-iT)^2+\rho^2]^{-\frac{1}{2}}+C[(a+iT)^2+\rho^2]^{-\frac{1}{2}}. \qquad (7)\end{aligned}$$

The quantity a is an approximate measure of the pulse width.*

Consider any value of the distance ρ, which is very large compared to a. Then (7) is large only when T is

* *Note added in proof.*—We are grateful to W. B. Bonnor for a reprint of his paper, J. Math. and Mech. 6, 203 (1957) where he considers this same wave form.

near $-\rho$ and near $+\rho$. At the one moment the imploding wave is going by; at the other moment it is reexploding towards infinity.

Let T be a large negative number, and ask for the shape of the pulse at this time. Introduce the dimensionless quantity x by the formula $\rho=-T+ax$. Then the pulse has the approximate form

$$\psi\sim 2C(-2aT)^{-\frac{1}{2}}\{[x+(1+x^2)^{\frac{1}{2}}]/2(1+x^2)\}^{\frac{1}{2}}, \quad (8)$$

indicated in Fig. 1. Pulses of other shapes can easily be constructed by combining the two terms in (7) with constant phase factors, $e^{i\delta}$ and $e^{-i\delta}$, respectively. The shape of the present pulse does not change with time until it comes close to the origin. The strength evidently grows inversely as the root of the distance of the peak of the pulse from the origin, as expected for a cylindrical wave. At the origin itself the dilatation factor ψ has the value

$$\psi=2Ca/(a^2+T^2). \quad (9)$$

The second dilatation factor γ is determined up to a constant by integrating (3) and (4), as follows:

$$\gamma=\tfrac{1}{2}C^2\{a^{-2}-\rho^2[(a-iT)^2+\rho^2]^{-2}-\rho^2[(a+iT)^2+\rho^2]^{-2}$$
$$-a^{-2}(T^2+a^2-\rho^2)[T^4+2T^2(a^2-\rho^2)+(a^2+\rho^2)]^{-\frac{1}{2}}\}. \quad (10)$$

Choose a fixed time T and an arbitrarily small quantity ϵ. Try to draw a sphere around the origin of a radius R so large that the dilatations, ψ and γ, are less than ϵ everywhere over the surface of this sphere. The task is an impossible one; there is no such sphere. The metric is everywhere regular, but it does not become asymptotically flat in the sense just mentioned.

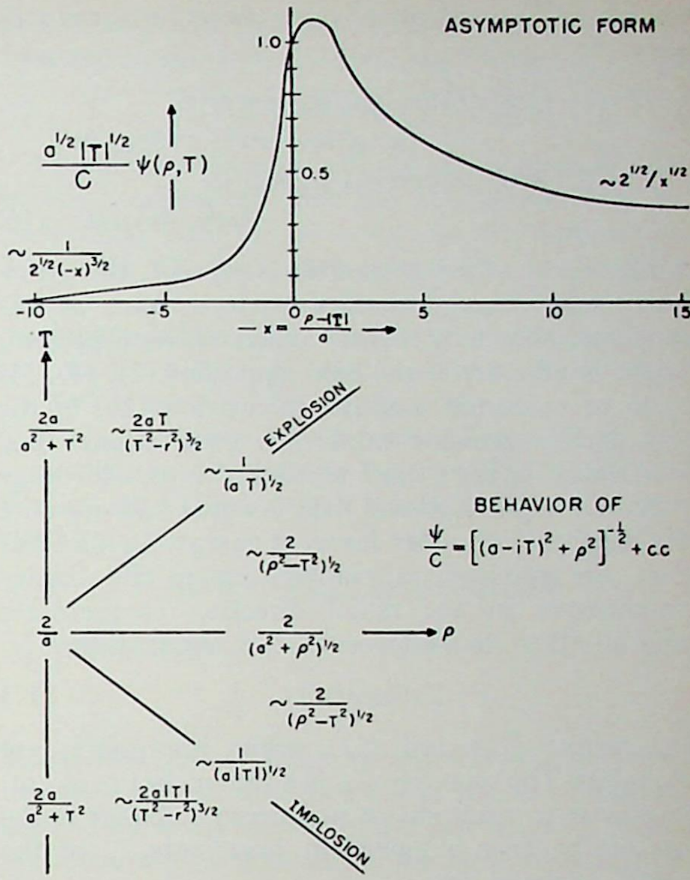


FIG. 1. The pulse-like cylindrical gravitational wave of Eq. (7) is concentrated near the radius $\rho=-T$ on the implosion run, for negative values of T; and concentrated near $\rho=+T$ when it reexpands out from the axis of symmetry. The upper diagram shows the pulse shape valid for both inward and outward runs when ρ and T are large compared with the pulse width, a. The lower diagram gives approximate representations of the pulse valid in the selected regions of the (ρ,T) diagram. The symbols r and ρ are used indiscriminately. The asymptotic value of the dilatation constant γ within the forward and backward light cone is taken to be zero, from which it follows that the asymptotic value of γ everywhere in the neutral zone is the constant

$$\gamma_{\text{asym}}=\int_0^\infty \rho(\psi_\rho{}^2+\psi_T{}^2)d\rho.$$

3. VANISHING OF THE PSEUDOTENSOR DENSITY OF ENERGY AND MOMENTUM

In evaluating the stress energy pseudotensor, we consider not only the monochromatic wave (5) and the pulse (7), not only the most general solution of the gravitational field equations (2)–(4), for empty space, but also the case where nongravitational energy is present. Then the field equations (2), (3), (4) no longer pertain. However, we limit ourselves to the case where the stress-energy tensor, T_{ik}, of this extra energy has cylindrical symmetry and where a metric of the form (1) still applies.

The pseudotensor, t_{ik}, of gravitational stress and energy is known not to be uniquely defined by the local conservation law, or requirement of zero divergence:

$$(-g)^{-\frac{1}{2}}(\partial/\partial x^\alpha)(-g)^{\frac{1}{2}}(T_k{}^\alpha+t_k{}^\alpha)=0 \quad (k=1,2,3,4). \quad (11)$$

We adopt the familiar Einstein choice for the stress energy pseudotensor, formulated by Tolman[4] in the following convenient language: (1) Define a Lagrange density function,

$$\mathfrak{L}=(-g)^{\frac{1}{2}}g^{\alpha\beta}(\Gamma_{\alpha\beta}{}^\mu\Gamma_{\mu\nu}{}^\nu-\Gamma_{\alpha\nu}{}^\mu\Gamma_{\beta\mu}{}^\nu). \quad (12)$$

(2) Define the quantity

$$s_i{}^{km}=(c^4/8\pi G)g^{k\lambda}\partial\mathfrak{L}/\partial g_{,m}{}^{i\lambda} \quad (13)$$

where $g_{,m}{}^{il}$ is an abbreviation for $\partial g^{il}/\partial x^m$. Then the divergence of this quantity gives an expression for the total stress energy pseudotensor,

$$T_i{}^k+t_i{}^k=(-g)^{-\frac{1}{2}}\partial s_i{}^{k\mu}/\partial x^\mu, \quad (14)$$

that satisfies the conservation law (11).

Calculation of (12), (13), and (14) is simplified by the diagonal character of the metric. We find

$$\partial\mathfrak{L}/\partial g_4{}^{44}=0;$$
$$\partial\mathfrak{L}/\partial g_1{}^{44}=\tfrac{1}{2}(-g)^{\frac{1}{2}}g^{11}g_{44}(g^{22}\partial g_{22}/\partial x^1+g^{33}\partial g_{33}/\partial x^1) \quad (15)$$

and thence the values of $s_4{}^{41}$ and $s_4{}^{44}$. We have no interest in the values of $s_4{}^{42}$ and $s_4{}^{43}$ because their derivatives with respect to $x^2=\theta$ and $x^3=\varphi$, respectively, are zero.

[4] R. C. Tolman, Phys. Rev. 35, 875 (1930); see C. Møller, *The Theory of Relativity* (Clarendon Press, Oxford, 1952) for a summary. Also, see especially L. Landau and E. Lifshitz, *The Classical Theory of Fields*, translated by M. Hamermesh (Addison-Wesley Press, Reading, Massachusetts, 1951), Chap. 11.

We end up with an identically zero value for the energy density,

$$T_4^4+t_4^4=(-g)^{-\frac{1}{2}}(c^4/16\pi G)(\partial/\partial x^1)(-g)^{\frac{1}{2}}g^{11}$$
$$\times(g^{22}\partial g_{22}/\partial x^1+g^{33}\partial g_{33}/\partial x^1)$$
$$=(c^4/8\pi G)(\partial/\partial\rho)(-1+2\rho\partial\psi/\partial\rho$$
$$-2\rho\partial\psi/\partial\rho)=0. \quad (16)$$

In other words, when no matter is present, the gravitational pseudoenergy density itself is zero, as has already been shown by Rosen.[2] When matter is present, the right-hand sides of the field equations (2), (3), (4) have to be corrected, but the calculation (16) of the energy density remains valid. The gravitational field automatically adjusts itself so that the pseudoenergy density of the gravitational field is equal and opposite to the density of all other forms of energy. In a similar fashion, the gravitational contribution to the density of momentum in the radial direction compensates exactly all other contributions to this momentum,

$$T_4^1+t_4^1=0. \quad (17)$$

The results (16) and (17), while impressive, are meaningless. The quantity T_{ik} is a tensor, but t_{ik} is not. A coordinate transformation produces no change in the values of the four independent invariants I_p of the tensor T_{ik}, defined by

$$\lambda^4+\lambda^3I_1+\lambda^2I_2+\lambda I_3+I_4 = \begin{vmatrix} (\lambda+T_1^1) & T_1^2 & T_1^3 & T_1^4 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ T_4^1 & T_4^2 & T_4^3 & (\lambda+T_4^4) \end{vmatrix}. \quad (18)$$

The corresponding quantities, i_1, i_2, i_3, i_4 for the pseudotensor t_{ik} are not invariant; they can be given arbitrary values at a point by suitable coordinate transformation or by suitably altering the definition (14) of the pseudoenergy tensor, consistently with the conservation law (11) or (3) by any combination of (1) and (2).

In physical terms, it has no well defined sense to say that the density of gravitational energy is zero, or that the flow of gravitational energy vanishes. No experiment could ever test whether these quantities are zero. The density of *other* forms of energy can be measured by the gravitational fields they produce. But gravitational field energy itself does not appear as a source term in Einstein's field equations, so there is no unambiguous way to speak of *localized* gravitational energy density as a source of a measurable gravitational field.

Only total energy has a well-defined significance, and even this only when very special conditions are met. There is no such quantity as total energy, for example, in a closed universe[4]; there the integrated conservation laws reduce to the trivial identity, zero equals zero. To define total energy it is necessary that the metric become asymptotically flat. More precisely, a coordinate system must exist in which the metric quantities go to the Euclidean flat space values at least as fast as 1/(distance from origin). Then the gravitational field falls off as fast as $1/r^2$ or faster, and a surface integral allows one to compute the mass or energy of the system unambiguously. This condition is not met in the case of an infinite cylindrical gravitational wave.†

The wave must be limited to a finite region of space to have a well-defined energy. Consider therefore a gravitational wave imploding, not on a straight line, but on a circular line. Let the thickness of the pulse a

† *Note added in proof.*—We are indebted to Professor M. Fierz for permission to quote from his letter to us of May 14, 1957.

"So we have $\gamma(\infty,T)=$constant, i.e., independent of T [from Eq. (4)]. Let $\psi(\rho,T)$ be a wave-packet fulfilling Eq. (2); we may assume that $\psi(\rho,T)=0$ for $\rho>R$—naturally this is not strictly true, but practically we may imagine ψ to have a finite spread. Now for $\rho\gg R$ we have

$$ds^2=e^{2\gamma}(d\rho^2-dT^2)+\rho^2d\varphi^2+dz^2,$$

where γ is a *positive constant*, namely the 'energy' in the ordinary sense of the wave field ψ:

$$\gamma(\infty,T)=\int_0^\infty \rho(\psi_\rho{}^2+\psi_T{}^2)d\rho=\text{const}$$

[where γ is normalized to zero at the origin]. Putting $\rho'=e^\gamma\rho$, $T'=e^\gamma T$, $\varphi'=e^{-\gamma}\varphi$, we have

$$ds^2=-dT'^2+d\rho'^2+\rho'^2d\varphi'^2+dz^2.$$

This looks pseudo-Euclidean. *But* the range of φ' is not $0\to2\pi$, but $0\to e^{-\gamma}2\pi$. This means that the geometry on a 'plane' $T=$const, $z=$const is the one on a *conical surface*. This holds as well as we want for any T if ρ (or ρ') is big enough. So, if we assume ψ to be regular everywhere the geometry is not 'regular' at infinity ($\rho\to\infty$), that is, it is not pseudo-Euclidean though the curvature vanishes for big ρ. Looking at the problem in this way, it is evident that the deviation from [Euclidean character] is not due to the coordinate system. This peculiar behavior is a property in the large."

This beautiful analysis of Fierz indicates that there is a subtle sense in which one can define an energy-like quantity per unit length of a cylindrically symmetrical gravitational wave, and indicates further that this energy-like quantity is positive definite. There is a most interesting analogy between the conical space of Fierz, and the curved space of the Schwarzschild metric, in the following sense: In both cases there is a departure from a Euclidean character, the integrated magnitude of which over a surface of radius ρ or r is independent of r and measures the total energy included within that surface. Of course, the area of the bounding surface increases faster with r in the second case than the first; correspondingly, the departure from the Euclidean character remains constant (as measured by $\Delta\varphi$) in the cylindrical case, and falls off with r in the spherical case. See also the discussion by Fierz at the end of reference 2.—We owe to Dr. Charles Misner an important additional remark about the definability of energy in a system that is not asymptotically Euclidean, and especially in a closed universe. Landau and Lifshitz show [reference 4, Eq. (11–88)] that the energy-momentum four vector can be expressed in the form of an integral over a closed 2-dimensional surface bounding a 3-dimensional space-like volume $P^\beta=\int h^{\beta0\alpha}dS^*_{0\alpha}$. When the three-dimensional space is closed, then the bounding surface S can be shrunk to zero. This circumstance might appear to argue, as indicated in the text, that the law of energy conservation reduces to the trivial identity, $0=0$. However, no closed surface can be covered without singularity by a single coordinate system. One has to use at least two coordinate patches. But $h^{\beta0\alpha}$ depends upon the gravitational stress-energy pseudotensor. Therefore its values in the two or more coordinate patches in the regions where they overlap are not related to each other as are the values of a true tensor. Consequently, it is *not* clear—Dr. Misner points out—that the energy integral will vanish. The remarks of Fierz and Misner raise the following question. Does there exist in relativity theory for a closed space a quite generally defined energy-like quantity that has not yet been clearly formulated?

be chosen very small compared to the radius, b, of the circle. For simplicity, let the initial conditions for the problem be formulated at the moment, $T=0$, of maximum implosion. The dilatation factors at this time we require to follow formulas like (7) and (10),

$$\psi \doteq 2C(a^2+\rho^2)^{-\frac{1}{2}}; \qquad \partial\psi/\partial T=0$$
$$\gamma \doteq (C^2/a^2)[(1+a^2/\rho^2)^{-2}-1]; \quad \partial\gamma/\partial T=0 \qquad (19)$$

provided that the distance ρ—now measured from the circular line—is small compared to the radius, b, of that line. We now have a toroidal gravitational wave instead of a cylindrical gravitational wave. For times before $T=0$ this wave implodes on the circle, and for later moments the wave explodes out from the circle. If the wave is weak enough ($C \ll a$), the familiar linearized theory of gravitational waves, or even Huygens' analysis, can be applied to discuss its behavior. It will approach asymptotically to a spherical wave pattern for very large values of the time. The wave will behave qualitatively in the same way when the dimensionless measure of strength C/a is larger, but then the analysis will be more complicated.

The point at issue is this: will the toroidal gravitational wave have any energy? By conservation laws the answer to this question at one time will answer it for all times. Therefore take the time $T=0$ of maximum concentration of the wave. The disturbance being confined to a limited region of space, one may assume the metric to be asymptotically flat. This permits definition of the total energy or mass of the gravitational wave. But is this energy unique? May it not be assigned at will by choosing arbitrary values for the coefficients of $1/r$ in the asymptotic formulas for the space part of the metric

$$ds^2 \sim (1+GM/2c^2r)^4(dx^2+dy^2+dz^2). \qquad (20)$$

Are not the initial conditions for the gravitational wave at the time $T=0$ a matter of arbitrary choice? Does the existence of this freedom mean that it has no sense to talk of gravitational energy and gravitational waves?

No such complete freedom exists.[5] Let $g_{\alpha\beta}$ $(\alpha,\beta=1,2,3)$ denote the space part of the metric at the time $T=0$, and let $R^{(3)}$ denote the curvature invariant of this three-dimensional space. Furthermore, introduce six quantities, $\varphi_{\alpha\beta}=\varphi_{\beta\alpha}$ $(\alpha,\beta=1,2,3)$, with the trace $\varphi=\varphi_\gamma{}^\gamma$, to measure the time rate of change of the metric,

$$\partial g_{\alpha\beta}/\partial T=\varphi_{\alpha\beta}. \qquad (21)$$

Then the three-dimensional "stretching rate tensor," $\varphi_{\alpha\beta}$ has to satisfy in the three space, $T=0$, the equations of Lichnerowicz and Fourès

$$(\varphi_\alpha{}^\beta-\delta_\alpha{}^\beta\varphi)_{;\beta}=0 \quad (\alpha=1,2,3)$$
$$\varphi^2-\varphi_{\alpha\beta}\varphi^{\alpha\beta}+R^{(3)}=0. \qquad (22)$$

At $T=0$, the first time derivative of all the metric quantities vanishes. It follows from (22) that the *three-dimensional curvature invariant of the metric $g_{\alpha\beta}$ must vanish everywhere* at the time $T=0$:

$$R^{(3)}=0 \quad \text{at} \quad T=0. \qquad (23)$$

This demand is automatically and accurately fulfilled by the line element (20). But this line element is the asymptotic form of the metric for the toroidal wave. Thus (23) is a differential equation for the metric that connects its asymptotic behavior (specified by the parameter M) with its behavior in the region of energy concentration, so *(23) provides a means to determine the "mass" or energy of the gravitational wave.*

The most general pure gravitational wave that is symmetric with respect to the time $T=0$ satisfies (23). A particular class of such solutions can be written in the form

$$ds^2=f^4(x,y,z)(dx^2+dy^2+dz^2), \qquad (24)$$

where the Laplacian of f must vanish:[6]

$$\nabla^2 f=0.$$

Apart from the trivial constant solution, all solutions of this equation diverge somewhere, as illustrated by (20). The everywhere regular solutions of (23) can therefore not be written in the conformal form (24). We have not so far found a scheme to construct and catalog all such regular solutions. The solution of this problem is central to further study of time-symmetric gravitational waves.

4. THE CYLINDRICAL GRAVITATIONAL WAVE PRODUCES A NONZERO CURVATURE

From now on we regard the cylindrical gravitational wave as the idealized limit of a toroidal gravitational wave, of finite and well-defined energy, when the radius b of the torus is extremely large in comparison with all other relevant physical dimensions.

To see that this wave is not a fictitious "coordinate" wave in a really flat space, it is enough to consider the Riemann curvature *tensor*, R_{ijkl}. Were the space really flat, this tensor would vanish in one coordinate system and therefore in all coordinate systems. However, by direct calculation from the metric (1) [without use of the field equations (2), (3), (4)] we find the components

TABLE I. Alteration in separation of nearby test particles as an appropriate invariant way to describe the polarization properties of the cylindrical gravitational wave.

Does separation at right change as determined by alteration in	Nature of separation of the two infinitesimal test particles		
	$\delta\rho$	δz	$\rho e^{-\psi}\delta\varphi$
Coordinates?	yes	no	$\delta\varphi$ unchanged
Invariant intervals?	yes	yes	yes

[5] See Y. Fourès-Bruhat, J. Rational Mech. Anal. 5, 951 (1956).

[6] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, 1926), p. 90, Eq. (28.7).

$$
\begin{aligned}
R_1{}^4{}_{14} &= \gamma_{TT}-\gamma_{\rho\rho}-\psi_{TT}+\psi_{\rho\rho}\\
R_1{}^2{}_{12} &= \frac{\gamma_\rho+\psi_\rho}{\rho}-\psi_\rho\gamma_\rho+\psi_{\rho\rho}-\psi_T\gamma_T+\psi_T{}^2\\
R_1{}^2{}_{42} &= -\psi_\rho\gamma_T+\psi_{T\rho}+\gamma_T/\rho-\psi_T\gamma_\rho+\psi_T\psi_\rho\\
R_1{}^3{}_{13} &= \psi_T\gamma_T-\psi_T{}^2+\psi_\rho\gamma_\rho-2\psi_\rho{}^2-\psi_{\rho\rho}\\
R_1{}^3{}_{43} &= \psi_T\gamma_\rho-3\psi_T\psi_\rho+\psi_\rho\gamma_T-\psi_{\rho T}\\
R_2{}^3{}_{23} &= \rho^{-2\gamma}\rho^2[\psi_\rho{}^2-\psi_T{}^2-\psi_\rho/\rho]\\
R_2{}^4{}_{42} &= \rho^{-2\gamma}\rho^2\left[\psi_{TT}-\psi_T\gamma_T+\frac{\gamma_\rho-\psi_\rho}{\rho}-\psi_\rho\gamma_\rho+\psi_\rho{}^2\right]\\
R_2{}^4{}_{12} &= \rho^{-2\gamma}\rho^2\left[\psi_{T\rho}-\psi_T\gamma_\rho+\psi_T/\rho+\frac{\gamma_T-\psi_T}{\rho}-\psi_\rho\gamma_T+\psi_\rho\psi_T\right]\\
R_3{}^4{}_{43} &= \rho^{4\psi-2\gamma}[\psi_T\gamma_T-2\psi_T{}^2-\psi_{TT}+\psi_\rho\gamma_\rho-\psi_\rho{}^2].
\end{aligned} \tag{25}
$$

Making use of the field equations (2), (3), (4) further simplifies these expressions, but does not reduce them all to zero. Existence of these nonzero components proves that the cylindrical gravitational waves propagate real curvature in space.

This departure from flatness produces real physical effects; the curvature modifies[7] the invariant distance between nearby infinitesimal test particles. Consequently objective existence must be attributed to the gravitational wave that produces these effects.

5. RESPONSE OF TEST PARTICLES TO THE CYLINDRICAL GRAVITATIONAL WAVE

An infinitesimal test particle follows the geodesic equation of motion

$$d^2x^i/ds^2+\Gamma_{\alpha\beta}{}^i(dx^\alpha/ds)(dx^\beta/ds)=0 \tag{26}$$

where the quantities $\Gamma_{jk}{}^i$ may be regarded as components of the gravitational field—dependent, of course, according to the equivalence principle, on the choice of the coordinate system. In the coordinate system (1) a test particle which is instantaneously at rest ($dx^1/ds=0$; $dx^4/ds=1$) experiences at that moment the acceleration

$$
\begin{aligned}
d^2\rho/dT^2 &= -\Gamma_{44}{}^1=\tfrac{1}{2}g^{11}\partial g_{44}/\partial x^1=-\partial\gamma/\partial\rho+\partial\psi/\partial\rho\\
&= C^2\rho\left\{\frac{\rho^2-(a-iT)^2}{[(a-iT)^2+\rho^2]^3}+\frac{\rho^2-(a+iT)^2}{[(a+iT)^2+\rho^2]^3}\right.\\
&\quad\left.+2\frac{\rho^2+a^2+T^2}{[\rho^2+(a-iT)^2]^{\frac{3}{2}}[\rho^2+(a+iT)^2]^{\frac{3}{2}}}\right\}\\
&\quad+C\rho\{[(a-iT)^2+\rho^2]^{-\frac{3}{2}}+[(a+iT)^2+\rho^2]^{-\frac{3}{2}}\}
\end{aligned} \tag{27}
$$

from the pulse type of gravitational wave. In particular, a particle once at rest at the axis of the cylinder remains forever at rest.

[7] F. Pirani, report at Chapel Hill conference (to be published).

To treat the motion of a particle that is some distance from the axis, we analyze the equation of motion to the first order of approximation: (1) we neglect terms in C^2, such as γ and its derivatives, in comparison to terms in C; (2) we evaluate the forces not at the new position, but at the old position; (3) we treat dx^4/ds as equal to 1; and (4) we write the pulse in the form

$$\psi \doteq f(\rho+T) \tag{28}$$

for negative values of T; and in the form

$$\psi = f(\rho-T) \tag{29}$$

for positive values of T. In this approximation the velocity has the value

$$
\begin{aligned}
d\rho/dT &\doteq \int_{-\infty}^{T}(d^2\rho/dT^2)dT\\
&\doteq \int_{-\infty}^{T}(\partial\psi/\partial T)dT=+\psi(\rho_0,T) \quad \text{for} \quad T<0\\
d\rho/dT &\doteq -\psi(\rho_0,T) \quad \text{for} \quad T>0.
\end{aligned} \tag{30}
$$

The particle experiences a sharp outward push (see Fig. 1) at about $T=-\rho$; then a weaker inward acceleration that brings the particle to rest at the time $T=0$. Then the inward slow acceleration resumes, the particle gains more and more velocity; and finally, at about $T=+\rho$, the particle experiences a final sharp outward acceleration that reduces it to rest. The outward migration of the particle up to the time $T=0$ is just balanced by the subsequent inward migration,

$$\Delta\rho=\int_{-\infty}^{0}(d\rho/dT)dT\doteq\int_{-\infty}^{0}\psi dT=C\pi. \tag{31}$$

The particle ends up where it started. A calculation to the first order that is really accurate does not replace $\partial\psi/\partial\rho$ by $\partial\psi/\partial T$ and gives for the migration, not (31), but

$$\Delta\rho=\int_{-\infty}^{0}\int_{-\infty}^{T'}(\partial\psi/\partial\rho)dT'dT=2C$$

for large ρ for the change from $T=-\infty$ to $T=0$.

The migration in coordinate gives no true measure of the change in distance between test particle and origin. To the first order this distance at the time $T=0$ is

$$
\begin{aligned}
\int(g_{11})^{\frac{1}{2}}d\rho &= \int_0^{\rho_0+\Delta\rho}e^{\gamma-\psi}d\rho\doteq\int_0^{\rho_0+\Delta\rho}(1-\psi)d\rho\\
&= \rho_0+\Delta\rho-\int_0^{\rho}\psi d\rho\\
&\doteq \rho_0+2C-2C\ln(2\rho/a).
\end{aligned} \tag{32}
$$

The particle moves, not farther from the origin, but

closer to it.‡ These results warn how dangerous it is to draw any conclusions about motions of test particles from changes in their *coordinates*. One must instead use invariant intervals between particles to measure the effect of the wave. We note (Table I) that the principal axes of the polarization ellipse lie in the z and φ directions, despite the fact that these coordinates do *not* change, while ρ does.

Can energy be extracted out of the gravitational wave? Yes, because the distance δz between nearby test particles changes with time. This change can be used to drive an engine. Where does the energy come from, it having been agreed that the pseudotensor of gravitational energy density and energy flow is zero in the chosen coordinate system? The maximum possible rate of absorption of energy by the engine will be governed by the rate at which the gravitational forces do work upon the test particles. These forces will be proportional to the first power of the mass of the typical test particle. The inward energy flux onto the test system ought therefore to be proportional to the first power of the mass of the test particle. Consider these quantities in the weak field approximation. Let Γ_{back} symbolize the gravitational field of the cylindrical wave. Likewise let Γ_{test} symbolize the gravitational field due to the moving test particles. Then the energy density is qualitatively of the form

$$\Gamma_{\text{back}}^2+2\Gamma_{\text{back}}\Gamma_{\text{test}}+\Gamma_{\text{test}}^2.$$

The first term, due to the background field, can be given arbitrary values by suitable choice of the coordinate system, and happens to vanish in the coordinate system (1). The middle term describes the flow of energy into the absorbing system.

Details of the energy flow require elaborate analysis. By analogy with the problem of an electromagnetic wave working on one or two test particles: (1) response of the particles, measured by changes in their invariant separation, is proportional to the first power of the field strength in the weak field approximation; (2) loss of energy from the driving wave to the test particle system is describable in terms of an interference between the primary wave and the secondary or scattered wave[8]; (3) when there is no true absorption in the test system, but only scattering, then the scattered energy, of second order in the field of the test particle, is the same in integrated value but is opposite in sign to the loss of energy from the primary wave, represented by the term linear in the field of the test particle; (4) the primary wave moves the test particle only transversely to the first order in the primary field strength, and to the second order in the primary field strength there is also a forward and backward response of the test particle.[9] A charged particle responding to an electromagnetic wave executes a figure eight motion—without however undergoing any net forward motion in this second approximation. Only when the radiative reaction of the secondary wave on the test particle is taken into account does one find a phase lag in the figure eight motion and a net forward impulse imparted to the charge. Radiation pressure comes in only when the scattering is taken into account. All these results are consistent with the conservation laws.

Similarities between gravitational and electromagnetic waves thus make it simple to draw a number of reasonable inferences. The significance of these inferences has a much more subtle character in the gravitational case than in the electromagnetic case. Neither field energy densities nor test particle motions have a meaning independent of the choice of coordinate systems. The simple observable consequences of wave action are instead *changes* in the separation of various nearby test particles—changes that are related to the covariant components of the curvature tensor R_{ijkl}, rather than to the values of the noncovariant field strengths, $\Gamma^i{}_{jk}$. This more complex character of the observation problem in gravitation physics is in full accord with Einstein's principle of equivalence.

This principle does not deny a physical reality to gravitational radiation but on the contrary leads to the only well defined way there is to express the influence of this radiation: in terms of its effect upon invariant space time intervals, such as the interval between two test bodies.

ACKNOWLEDGMENTS

We express our appreciation to Professor S. R. De Groot for the hospitality of the Lorentz Institute of the University of Leiden where most of this work was done. We acknowledge stimulating discussions with Dr. C. W. Misner, Mr. Peter Putnam, Professor Marcel Reisz, and Professor Peter G. Bergmann. Both of us express gratitude for John Simon Guggenheim Memorial Foundation Fellowships; in addition one of us (Joseph Weber) is indebted to the National Research Council for a Merck Fellowship and the other (John Wheeler) is indebted to the United States (Fulbright) Educational Foundation in the Netherlands for a travel grant.

‡ *Note added in proof.*—To the second order the velocity of the test particle originally at ρ_0 follows from integration of

$$d^2\rho/dT^2 \doteqdot (\partial\psi/\partial\rho)_\rho+(d\rho/dT)(\partial\psi/\partial T)_{\rho_0}-(\partial\gamma/\partial\rho)_{\rho_0}.$$

Here it is legitimate in the same order to write

$$(\partial\psi/\partial\rho)_\rho \doteqdot \psi_0'+(\rho-\rho_0)\psi_0''.$$

The integration gives for the velocity at the time $T=+\infty$ the values $-2C^2/a^2$ and $-2\pi C^2\rho_0/a^3$ for values of ρ_0 large and small compared to a, respectively. This finite velocity produces in the course of an infinite time an infinite displacement. This circumstance does not affect the validity of the expansion in powers of the wave strength, C.

[8] See Bohr, Peierls, and Placzek, Nature **144**, 200 (1939); N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon Press, Oxford, 1949), second edition, Chap. VIII; and N. F. Mott, Proc. Roy. Soc. (London) **A133**, 228 (1931).

[9] We are indebted to Professor Robert Thompson for illuminating discussions on this point a number of years ago.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 3 JULY, 1957

On the Question of a Neutrino Analog to Electric Charge

JOHN KLAUDER AND JOHN A. WHEELER

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

1. INTRODUCTION

THE Maxwell equations apply[1] in a multiply connected space as well as in a simply connected space; the "handles" or "worm holes" in a multiply connected space trap lines of force, in such a way that one can associate an integrated flux of lines of force, or "charge," with each "worm hole"; and the charge so defined stays constant in time as a consequence of Maxwell's equations, regardless of the complication of the initial field values, and regardless of the way the metric changes with time, as long as the topology does not change. The theorem that the number of linearly independent constants of charge is equal to the number of worm holes (or to the second Betti number, R_2, of the 3-space) is a consequence of the well known theory of harmonic forms.[2]

This note enquires whether the two-component Dirac equation for the neutrino in curved space[3] resembles the Maxwell equations in the following sense: *does the two-component neutrino equation admit zero frequency solutions in a multiply-connected time-independent metric?* If such solutions, $\psi_1, \psi_2, \cdots$, exist, then one can construct a general solution of the following form

$$\psi = Q_1\psi_1 + Q_2\psi_2 + \cdots,$$

where the Q_k's are constants. One might then ascribe to the constants, Q_k, the name of "neutrino charge" and look for conservation properties—when the metric is allowed to vary—analogous to the conservation already proven for electric flux.

The simplest multiply-connected space has the topology of a "torus" defined by the metric in (5) and (6). This space admits a zero-frequency solution of Maxwell's equations.[4] In the same space the two-component neutrino wave equation is treated by the method of the separation of variables and it is proved that there exists no nonvanishing solution of this equation that is acceptable according to present standards. Consequently, *there is no reason to expect for the neutrino field a concept of charge having any direct analogy to electric charge.*

2. TWO-COMPONENT EQUATION IN CURVED SPACE-TIME SOLUTION BY SEPARATION OF VARIABLES

Unfortunately, no theory of harmonic spinor fields exists, on which one can draw, as one can on the existing highly developed theory of harmonic vector and scalar fields. There is no harmonic *scalar* function in a closed three space which is not a trivial constant. Does the spinor neutrino field more resemble, in respect to number of independent solutions, the scalar field ($N=0$) or the vector field ($N=R_2$)? To prove that the number of nontrivial spinor fields is not R_2 as for a vector field, it is sufficient to give one contrary example. This example is the subject of the remainder of this note.

The two component specialization of the Dirac equation takes the form[5]

$$\sigma^k{}_{\dot{\mu}\rho}[\partial\psi^\rho/\partial x^k + \Gamma^\rho{}_{\alpha k}\psi^\alpha] = 0. \tag{1}$$

Latin indices, such as k, range from 1 to 4; and Greek indices range from 1 to 2, thus $\rho=(1,2)$. Dotted indices denote complex conjugation and are summed only against other dotted indices. The quantities $\sigma^k{}_{\dot{\mu}\rho}$ are defined by the relation

$$\sigma^{k\tau\dot{\mu}}\sigma^l{}_{\dot{\mu}\rho} + \sigma^{l\tau\dot{\mu}}\sigma^k{}_{\dot{\mu}\rho} = 2g^{kl}\delta_\rho{}^\tau \tag{2}$$

and the metric is chosen to have the signature $(1,1,1,-1)$. Lastly, four 2×2 *connection matrices*, $\Gamma^\rho{}_{\alpha k}$, are defined up to an additive multiple of the unit matrix by the following equation

$$\partial\sigma^k{}_{\dot{\mu}\rho}/\partial x^s + \Gamma^k{}_{rs}\sigma^r{}_{\dot{\mu}\rho} - \Gamma^\tau{}_{\rho s}\sigma^k{}_{\dot{\mu}\tau} - \Gamma^{\dot{\sigma}}{}_{\dot{\mu}s}\sigma^k{}_{\dot{\sigma}\rho} = 0, \tag{3}$$

where the $\Gamma^k{}_{rs}$ are the usual Christoffel symbols.

Restricting attention to a special class of spaces, consider an Euclidean 4-space (x,y,z,w) where the fourth-dimension w has nothing to do with time. In this flat 4-space pick out a curved three-dimensional hypersurface, or 3-space, by way of

$$[(x^2+y^2+z^2)^{\frac{1}{2}} - \tau]^2 + w^2/\rho^2 = 1. \tag{4}$$

---

[1] J. A. Wheeler, Phys. Rev. **97**, 511 (1955); C. W. Misner and J. A. Wheeler (to be published).

[2] A. Lichnerowicz, *Algèbre et analyse linéaires* (Masson et Cie Paris, 1956), second edition; G. de Rham, *Variétés différentiables: formes, courants, formes harmoniques* (Hermann et Cie, Paris, 1955); W. V. D. Hodge, *The Theory and applications of Harmonic Integrals* (Cambridge University Press, Cambridge, 1952), second edition.

[3] See, for example, D. R. Brill and J. A. Wheeler, Revs .Modern Phys. **29**, 465 (1957).

[4] For zero frequency these equations reduce to two uncoupled pairs of equations, one pair for the electric field, the other for the magnetic field. It is sufficient to consider the first pair: div $\mathbf{E}=0$, curl $\mathbf{E}=0$. There is *no* solution of these equations in a spherical closed space, but in a toroidal space of the form (5) and (6), there exists one nonzero solution in which the lines of force encircle the torus in the χ direction. This is one of the simplest examples of a harmonic vector field. *Note added in proof.*—This 3-space is $S^1\times S^2$ while what is usually called a torus is $T^3=S^1\times T^2=S^1\times S^1\times S^1$.

[5] We use here nearly the same spinor notation as that used by W. L. Bade and H. Jehle, Revs. Modern Phys. **25**, 714 (1953).

Here τ and ρ are parameters defining a particular space. For example, a "toroidal" space would have $\tau>1$ and $\rho>0$, while a spherical space results for $\tau=0$, $\rho=1$. Points in the three-space can be specified by the three coordinates x, y, z. or more conveniently by three orthogonal angle variables, χ, θ, φ, the latter two being analogous to the usual angles of a spherical coordinate system.

Now introduce the time variable, or "cotime," $T=ct$. The four-dimensional metric becomes

$$ds^2=A(\chi)d\chi^2+B(\chi)[d\theta^2+\sin^2\theta d\varphi^2]-dT^2, \quad (5)$$

where

$$A(\chi)=\cos^2\chi+\rho^2\sin^2\chi$$
$$B(\chi)=(\tau+\sin\chi)^2. \quad (6)$$

To satisfy (2) the 2×2 spin matrices, $\sigma^k{}_{\dot{\mu}\rho}$ are selected to have the values

$$\sigma^1{}_{\dot{\mu}\rho}=A^{-\frac{1}{2}}\sigma_x; \qquad \sigma^2{}_{\dot{\mu}\rho}=B^{-\frac{1}{2}}\sigma_y;$$
$$\sigma^3{}_{\dot{\mu}\rho}=B^{-\frac{1}{2}}\sigma_z/\sin\theta; \quad \sigma^4{}_{\dot{\mu}\rho}=1 \quad (7)$$

in terms of the standard Pauli matrices,

$$\sigma_x=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}; \quad \sigma_y=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}; \quad \sigma_z=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \quad (8)$$

According to Bade and Jehle[5] the 2×2 matrices with raised spin indices have the values

$$\sigma^{1\dot{\rho}\sigma}=A^{-\frac{1}{2}}\sigma_x; \qquad \sigma^{2\dot{\rho}\sigma}=-B^{-\frac{1}{2}}\sigma_y;$$
$$\sigma^{3\dot{\rho}\sigma}=B^{-\frac{1}{2}}\sigma_z/\sin\theta; \quad \sigma^{4\dot{\rho}\sigma}=-1. \quad (9)$$

From the metric (5) the Christoffel symbols, $\Gamma^k{}_{rs}$, may be computed. Since the neutrino has no charge the usual electromagnetic fields are excluded in the following determination of the connective matrices, $\Gamma^\tau{}_{\rho k}$:

$$\Gamma^\tau{}_{\rho 1}=\Gamma^\tau{}_{\rho 4}=0$$
$$\Gamma^\tau{}_{\rho 2}=-i(AB)^{-\frac{1}{2}}B'\sigma_z/4 \quad (10)$$
$$\Gamma^\tau{}_{\rho 3}=i(AB)^{-\frac{1}{2}}\sin\theta B'\sigma_y/4-i\cos\theta\sigma_x/2.$$

Here B' is an abbreviation for $dB/d\chi$. Substituting into (1) we obtain the first-order two-component neutrino wave equation

$$A^{-\frac{1}{2}}\sigma_x(\partial/\partial\chi)B^{\frac{1}{2}}\psi+(\sin\theta)^{-\frac{1}{2}}\sigma_y(\partial/\partial\theta)(\sin\theta)^{\frac{1}{2}}\psi$$
$$+(\sin\theta)^{-1}\sigma_z\partial\psi/2\varphi+\partial\psi/\partial T=0. \quad (11)$$

Our interest centers around harmonic or zero-frequency solutions; hence the last term in (11) will be put equal to zero. The solution of (11) proceeds by assuming a space dependence of the form

$$\psi=B^{-\frac{1}{2}}\begin{pmatrix}f(\theta)\\ g(\theta)\end{pmatrix}\exp\left(k\int_{\chi_0}^{\chi}(A/B)^{\frac{1}{2}}d\chi+im\varphi\right). \quad (12)$$

Substituting into (11) and multiplying by σ_y the "radial" or χ dependent part of ψ comes out as a common factor, leaving only an equation in the angle θ:

$$[(\sin\theta)^{-\frac{1}{2}}(d/d\theta)(\sin\theta)^{\frac{1}{2}}-ik\sigma_z-(m\sigma_x/\sin\theta)]\begin{pmatrix}f\\ g\end{pmatrix}=0 \quad (13)$$

equivalent to the two equations

$$(\sin\theta)^{-\frac{1}{2}}(d/d\theta)(\sin\theta)^{\frac{1}{2}}f-ikf-(mg/\sin\theta)=0$$
$$(\sin\theta)^{-\frac{1}{2}}(d/d\theta)(\sin\theta)^{\frac{1}{2}}g+ikg-(mf/\sin\theta)=0. \quad (14)$$

The radial factor in (12) is only acceptable if the number k has one or another purely imaginary value or is zero, whereas the angular factor is only acceptable if k is a real nonzero integer. No acceptable solution meets both requirements.

The radial factor has the value

$$\psi_\chi=B^{-\frac{1}{2}}\exp\left(k\int_{\chi_0}^{\chi}(A/B)^{\frac{1}{2}}d\chi\right). \quad (15)$$

This radial factor should be nonsingular, and should be periodic with period 2π in χ. Therefore k must have the purely imaginary or zero value

$$k=(2\pi i)\cdot(\text{integer})\Big/\int_0^{2\pi}(A/B)^{\frac{1}{2}}d\chi, \quad (16)$$

where the integrand never changes sign and never even goes to zero.

Equations for the angular part of ψ similar to (14) have been investigated by Schrödinger.[6] They differ from ours because of our choice of representation of the matrices, $\sigma^k{}_{\dot{\mu}\rho}$. In no other respect but this formal one do the angular equations differ from the familiar angular equations for the Dirac electron. General relativity effects put in no appearance. Only those states meet present standards of acceptability for which k is a nonzero integer,[7] and for which m has one of the values

$$m=-|k|-\tfrac{1}{2},\ -|k|-\tfrac{3}{2},\ \cdots,\ |k|-\tfrac{3}{2},\ |k|-\tfrac{1}{2}.$$

The value $k=0$ is not among the allowed values. Consequently *there exists no harmonic spinor field in the given metric.*

ACKNOWLEDGMENTS

We wish to thank Professor V. Bargmann for his help and Charles W. Misner for a number of illuminating discussions.

---

[6] E. Schrödinger, Commentationes Pontif. Acad. Sci. **2**, 321 (1938).

[7] For a table of values of k and corresponding values of j and l, see for example L. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1940), first edition, p. 322. We investigated the two independent solutions for $k=0$, $m=0$,

$$\begin{pmatrix}f\\ g\end{pmatrix}=(\sin\theta)^{-\frac{1}{2}}\begin{pmatrix}1\\ 0\end{pmatrix} \quad \text{and} \quad (\sin\theta)^{-\frac{1}{2}}\begin{pmatrix}0\\ 1\end{pmatrix}.$$

Both are square integrable. However, operation on either by the angular momentum operators that transform $\psi(k,m)$ to $\psi(k,m+1)$ or $\psi(k,m-1)$ *ought* to give a zero result but instead give divergent functions. In other words, the angular momentum operators become non-Hermitian operators if the state $k=0$, $m=0$ is accepted.

Quantization of the Interacting Fields of Electrons, Electromagnetism, and Gravity*†

F. J. BELINFANTE, D. I. CAPLAN, AND W. L. KENNEDY

Purdue University, Lafayette, Indiana

CONTENTS

1. INTRODUCTION

SOME of the progress made during the past few years by others as well as by ourselves in the theory of quantization of the gravitational field in interaction with the Maxwell field and the Dirac electron field is reviewed. The reason why we consider the latter two fields in interaction with gravity is that the theory of these fields is representative for the complications due to gauge invariance as well as for the complications arising from the spinor character and the Fermi-Dirac quantization of a fermion field. Once these difficulties will have been solved for this representative case, we expect that quantization of other fields interacting with the gravitational field will not meet serious obstacles.

For simplicity we have assumed a simple topology of the universe, similar to that of special relativity theory. We also assume that all fields drop off to zero sufficiently strongly at spatial infinity for existence of a free energy-momentum four-vector.[1] Our coordinate system then may be chosen to approach an orthogonal Lorentz frame at spatial infinity. Matter and radiation fields then should be considered as occupying some finite volume in infinite space.

These assumptions may be unjustifiable, as space time might have a "Swiss-cheese" structure along world lines of elementary particles[2] or a foamy structure full of "worm holes" corresponding to fluctuations in the vacuum field,[3] or space may be closed.[4] Such possibilities deserve careful investigation, as they may lead to important consequences. Some of the conclusions arrived at here might be invalidated if the universe were full of holes, loops, or knots.

We commence by considering what kind of "classical theory" may serve as a starting point for inventing a quantum theory. This leads to a "modernization" of the classical theory of fermions, replacing the c-number formulation of the classical theory by a "quasi-c-number" formulation. This necessitates a redefinition of the rules of differentiation, of Poisson brackets, etc. In quantizing such a classical theory we use some of the devices proposed or suggested by Dirac,[5] such as "weakly vanishing quantities" and "modified Poisson brackets," altering and modernizing these methods as needed. It then turns out that the canonical variables derived from the original Lagrangian do not satisfy canonical commutation relations. New variables are introduced which do satisfy canonical commutation relations.[6] We also indicate a method by which one

* Supported in part by the National Science Foundation.
† Partially based upon parts of theses submitted by the second and third authors in partial fulfillment of the requirements for a Ph.D. degree at Purdue University.

[1] See, for instance, C. Møller, *The Theory of Relativity* (Clarendon Press, Oxford, 1952), pp. 339–340, or H. Weyl, *Space-Time-Matter* (Dover Publications, New York, 1950), pp. 269–272.
[2] F. J. Belinfante, Phys. Rev. **98**, 793 (1955).
[3] J. A. Wheeler and C. Misner, Chapel Hill Conference (1957). See Science News Letter **71**, 67 (February 2, 1957).
[4] See, for instance, H. Weyl, *Space-Time-Matter* (Dover Publications, New York, 1950), p. 278.
[5] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
[6] B. S. DeWitt and C. M. DeWitt, Phys. Rev. **87**, 116 (1952).

can arrive at this canonical quantization without use of Dirac's modified Poisson brackets.[7,8]

The proof of covariance of this quantization procedure meets difficulties partly due to vagueness of the meaning of the weakly vanishing quantities of Dirac's method. These difficulties are avoided in the altered theory, discussed in part II.‡ The altered theory is a theory which has been altered by adding noncovariant elements. Thus one avoids the existence of "constraints," and replaces them by "auxiliary conditions." The latter prevent the added terms in the Lagrangian from having any effect on predictions of the theory. The altered theory may well serve as a means for dealing with applications of the quantum theory of gravitation, just as most predictions of quantum electrodynamics were obtained from the similarly altered theory of Fermi further developed by Tomonaga, Schwinger, Feynman, Dyson, and others, and not by the "true" gauge-independent theory originated by Pauli and further developed by Belinfante.

The problem remains as to what questions one is allowed to ask the theory. This problem is closely related to the "initial-value problem" of the classical theory, and to the problem of determining the "true variables" of the gravitational theory. This is briefly introduced in Part III. Another problem introduced without a definite solution is that of whether there exists in the general theory of relativity such a thing as the "interaction representation" of Lorentz-covariant field theory.

PART I. QUANTIZATION OF THE "CLASSICAL" THEORY

2. MODERNIZATION

Quantum theories usually are derived from classical theories by postulating that the commutators of the variables no longer all vanish but have given values. By the classical theory of electrons one then means the wave-mechanical theory of Dirac.

Classical theories of Einstein-Bose fields are characterized by the fact that they are described by c-number fields, that is, by ordinary commutative algebraic quantities. The classical theory owes its limited validity for phenomena involving many quanta to the fact that it is the limiting case of the quantum-mechanical theory for fields of which the squares or products have expectation values much larger than the elementary quantum. It is the limiting case of the quantum-mechanical theory in the hypothetical limit in which Planck's constant h is made much smaller than any product of fields; that is, the limit for $h\rightarrow 0$, provided that this limit is taken in the correct way. In particular, the commutation relations[9]

$$[q_i(\mathrm{x},t);\, q_j(\mathrm{x}',t)] \equiv 0, \qquad (2.1a)$$

$$[p^i(\mathrm{x},t);\, p^j(\mathrm{x}',t)] \equiv 0, \qquad (2.1b)$$

$$[q_i(\mathrm{x},t);\, p^j(\mathrm{x}',t)] \equiv i\hbar c\delta_i{}^j\, \delta_3(\mathrm{x}-\mathrm{x}') \qquad (2.1c)$$

in this limit become

$$[q_i(\mathrm{x},t);\, q_j(\mathrm{x}',t)] = 0, \qquad (2.2a)$$

$$[p^i(\mathrm{x},t);\, p^j(\mathrm{x}',t)] = 0, \qquad (2.2b)$$

$$[q_i(\mathrm{x},t);\, p^j(\mathrm{x}',t)] = 0, \qquad (2.2c)$$

so the field variables describing Einstein-Bose fields in the classical limit all become commutative quantities, which justifies their treatment as c numbers.

It is seen from (2.2) that there is no similar justification for describing Fermi-Dirac fields in classical approximation by c numbers. For Fermi-Dirac fields, the "commutators" [;] are *anti*commutators, and (2.2) show that in "classical approximation" $\hbar\rightarrow 0$ Fermi-Dirac fields should be described by *purely anticommutative* quantities, which might be described by ordinary c numbers but for the fact that they flip sign when commuted with another such quantity. We call such quantities *quasi-c-numbers*. "Classical" theories of fermions in which ordinary c numbers would have been substituted for quasi-c-numbers are no longer the limiting case of the quantum theory and therefore may lead to difficulties.

The amendment of the classical theory for taking care of the anticommutativity of all "Fermi-Dirac quantities" among each other we shall call the *modernization* of the classical theory. By a *Fermi-Dirac quantity* we mean an expression in which each term contains an *odd* number of Fermi-Dirac field variables as factors while an expression in which each term contains an *even* number of Fermi-Dirac field variables (irrespective of what fermion fields those variables describe) will be called an *Einstein-Bose quantity*.

If Ψ is a state in which the quantity q has the value q', then Ψ is an eigenfunction of q belonging to the eigenvalue q':

$$q\Psi = q'\Psi. \qquad (2.3)$$

For Einstein-Bose fields, q may be an observable and q' an ordinary c number. For a Fermi-Dirac field variable q we would call Ψ an eigenfunction of q if (2.3) would be valid with q' a quasi-c-number. The main difference between an operator as q and a quasi-c-number such as q' is that the anticommutator of q with p would be a delta function, while the anticommutator of q' with p would vanish. In this modernized fashion, there are now simultaneous eigenfunctions of

[7] F. J. Belinfante and J. C. Swihart, "A theory of gravitation and its quantization," a report of research supported by the National Science Foundation and the Purdue Research Foundation (1954).

[8] F. J. Belinfante and J. C. Swihart, "Phenomenological linear theory of gravitation," Part III [Ann. Phys. 2 (1957)], and Part IV (to be published later).

‡ The word "altered" is used to avoid confusion with the use of "modified" for denoting the change of the Poisson brackets that was proposed by Dirac.

[9] We define p as $\partial L/\partial q_{,0}$ with $q_{,0}\equiv\partial q/\partial x^0\equiv\dot{q}/c$. By $\equiv$ we denote q-number equalities.

a Fermi-Dirac field variable at all points of space, just as there are for Einstein-Bose field variables. The old objection that the Fermi-Dirac field q at two different points 1 and 2 do not even commute and therefore have no simultaneous eigenfunction no longer holds, since the anticommutativity of the operators q_1 and q_2 is now exactly what is needed for the existence of a simultaneous eigenfunction, according to

$$q_1q_2\Psi = q_1(q_2'\Psi) = q_1'\,q_2'\Psi = -\,q_2'\,q_1'\Psi = -\,q_2(q_1'\Psi) = -\,q_2q_1\Psi. \quad (2.4)$$

Thus, quasi-c-numbers are useful not only in the "classical" theory of fermion fields, but also in the quantum theory of such fields, making it more similar to the theory of boson fields.

In theories of interacting fermion and boson fields, the "commutator" of two quantities A and B' is defined by

$$[A\,;B'] \equiv AB' - (-)^{u_Au_B}\,B'A, \quad (2.5)$$

where $u_A = 1$ for a Fermi-Dirac quantity A, while $u_A = 0$ for an Einstein-Bose quantity A. In the classical theory, such "commutators" vanish, thus making *any* pair of Fermi-Dirac quantities anticommutative. This lack of ordinary commutativity of field variables even in the classical limit necessitates a distinction between differentiation from the front (the left) and differentation from the back (from the right) of Einstein-Bose quantities with respect to Fermi-Dirac variables, in the classical as well as in the quantum theory. These two types of derivatives have been called the *ante*derivative and the *post*derivative[10,11]:

$$\partial^AF/\partial q \equiv (\delta q)^{-1}(\delta F); \quad \partial^PF/\partial q \equiv (\delta F)(\delta q)^{-1}. \quad (2.6)$$

Here, the variation of a quantity is assumed to be a quasi-c-number of the same (Einstein-Bose or Fermi-Dirac) character as the quantity varied. According to this definition, the derivative of a relation like (2.1c) with respect to q from the left even for a fermion field gives the identity $p - p = 0$, instead of the embarrassing relation $p + p = 0$ to which old-fashioned differentiation would give rise.

On the basis of the definitions (2.6), the modernized rules of differentiation are easliy set up.[10] One finds

$$(\partial^AQ_1/\partial q_0) \equiv (-)^{u_0(u_1+1)}(\partial^PQ_1/\partial q_0)\,; \quad (2.7\text{a})$$

$$(\partial^A/\partial q_0)(Q_1Q_2) \equiv (\partial^AQ_1/\partial q_0)Q_2 + (-)^{u_0u_1}\,Q_1(\partial^AQ_2/\partial q_0)\,; \quad (2.7\text{b})$$

$$(\partial^A/\partial q_1)(\partial^A/\partial q_2)Q \equiv (-)^{u_1u_2}(\partial^A/\partial q_2) \times(\partial^A/\partial q_1)Q, \quad (2.7\text{c})$$

$$(\partial^A/\partial q_1)(\partial^P/\partial q_2)Q \equiv (\partial^P/\partial q_2)(\partial^A/\partial q_1)Q, \quad (2.7\text{d})$$

and so on. The relations

$$(\partial^A/\partial q)F[Q_1(q),Q_2(q)\cdots] = \textstyle\sum_{\rm i}(\partial^AQ_{\rm i}/\partial q)(\partial^AF/\partial Q_{\rm i}), \quad (2.8\text{a})$$

$$(\partial^P/\partial q)F[Q_1(q),Q_2(q)\cdots] = \textstyle\sum_{\rm i}(\partial^PF/\partial Q_{\rm i})(\partial^PQ_{\rm i}/\partial q) \quad (2.8\text{b})$$

are valid in the modernized classical theory but in general are not valid in quantum theory unless one understands the right-hand members as meaning that $(\partial Q_{\rm i}/\partial q)$ is to be inserted in F at the place where $Q_{\rm i}$ is differentiated away in $\partial F/\partial Q_{\rm i}$. Otherwise, trouble arises from the fact that $\delta Q_{\rm i} = (\partial^PQ_{\rm i}/\partial q)\delta q$, contrary to δq itself, is not a quasi-c-number and may neither commute nor anticommute with some other factors in F.

Modernization will also affect the definition of the Poisson brackets. The reason why in quantum mechanics we are interested in Poisson brackets (P.B.) is their relation to the commutators, supposedly according to

$$\lim_{\hbar\to 0}\frac{[A\,;B']}{i\hbar c} = (A,B'). \quad (2.9)$$

Here, according to (2.5), the left-hand member changes sign under interchange of the quantities A and B', except if both are Fermi-Dirac quantities, in which case interchange leaves the "commutator" (2.5) invariant. Therefore, in the modernized classical theory the P.B. should be defined in such a way that the right-hand member of (2.9) has the same properties under interchange of A and B'. For functions of the q and p and their spatial derivatives $q_{,r} \equiv \partial q/dx^r$ and $p_{,r} \equiv \partial p/\partial x^r$ at two simultaneous points $(\mathbf{x},x^0)$ and $(\mathbf{x}',x^0)$, this leads us to define the modernized P.B. of A and of B' by

$$\begin{aligned}(A(\mathbf{x},x^0),\,B(\mathbf{x}',x^0)) \equiv \textstyle\sum_{\rm i}\Bigg\{&\left[\frac{\partial^PA}{\partial q_{\rm i}}+\frac{\partial^PA}{\partial q_{{\rm i},r}}\frac{\partial}{\partial x^r}\right]\left[\frac{\partial^AB'}{\partial p'^{\rm i}}+\frac{\partial^AB'}{\partial p'^{\rm i}{}_{,s'}}\frac{\partial}{\partial x'^s}\right]\\ &-(-)^{u_Au_B}\left[\frac{\partial^PB'}{\partial q'_{\rm i}}+\frac{\partial^PB'}{\partial q'_{{\rm i},s}}\frac{\partial}{\partial x'^s}\right]\\ &\times\left[\frac{\partial^AA}{\partial p^{\rm i}}+\frac{\partial^AA}{\partial p^{\rm i}{}_{,r}}\frac{\partial}{\partial x^r}\right]\Bigg\}\delta_3(\mathbf{x}-\mathbf{x}'). \quad (2.10)\end{aligned}$$

These P.B. of the modernized classical theory satisfy the identities

$$(A,B') = -\,(-)^{u_Au_B}(B',A), \quad (2.11)$$

$$(-)^{u_Au_C}(A,(B',C'')) + (-)^{u_Bu_A}(B',(C'',A)) + (-)^{u_Cu_B}(C'',(A,B')) = 0, \quad (2.12\text{a})$$

$$(-)^{u_Bu_A}((B',C''),A) + (-)^{u_Cu_B}((C'',A),B') + (-)^{u_Au_C}((A,B'),C'')=0, \quad (2.12\text{b})$$

$$(A',f[B_1,B_2,\cdots]) = \textstyle\sum_{\rm j}(A',B_{\rm j})(\partial^Af/\partial B_{\rm j}), \quad (2.13\text{a})$$

$$(f[B_1,B_2,\cdots],A') = \textstyle\sum_{\rm j}(\partial^Pf/\partial B_{\rm j})(B_{\rm j},A'). \quad (2.13\text{b})$$

[10] F. J. Belinfante, Physica **7**, 765 (1940).

[11] J. Schwinger [Phys. Rev. **91**, 713 (1953)] calls these derivatives the left and the right derivative.

The relations (2.11) and (2.12) match similar identities for the commutators (2.5), while the left-hand and right-hand members of each Eq. (2.13) with the P.B. replaced by commutators (2.5) in the quantum theory need not be equal except in the classical limit $\hbar \to 0$ in which all factors become purely commutative or anti-commutative. [Compare the note below (2.8).]

In a representation in which the state vector Ψ according to (2.3) is a modernized eigenfunctional of the fields $q(\mathbf{x})$ for all $\mathbf{x}$ at a given time x^0, the operators $p(\mathbf{x})$ satisfying the commutation relations (2.1) may be represented by the variational operators

$$p^j(\mathbf{x}) = (-)^{u j}(-i\hbar c)\{\mathfrak{d}^A/\mathfrak{d}q_j(\mathbf{x})\}. \qquad (2.14)$$

The meaning of the variational quotient here is given by

$$\mathfrak{d}^A q(\mathbf{x}')/\mathfrak{d}q(\mathbf{x}) = \delta_3(\mathbf{x}'-\mathbf{x}), \qquad (2.15a)$$

so that, with $f(\mathbf{x}')=f[q(\mathbf{x}'), q_{,s}(\mathbf{x}')]$ and

$$\mathfrak{f} = \int f(\mathbf{x}')d^3\mathbf{x}',$$

$$\frac{\mathfrak{d}^A f(\mathbf{x}')}{\mathfrak{d}q(\mathbf{x})} = \left[\frac{\partial^A f(\mathbf{x}')}{\partial q(\mathbf{x}')} + \frac{{}^\bullet\partial^A f(\mathbf{x}')}{\partial q_{,s}(\mathbf{x}')}\frac{\partial}{\partial x'^s}\right]\delta_3(\mathbf{x}'-\mathbf{x}), \qquad (2.15b)$$

$$\frac{\mathfrak{d}^A F(\mathfrak{f})}{\mathfrak{d}q(\mathbf{x})} = \frac{\mathfrak{d}^A \mathfrak{f}}{\mathfrak{d}q(\mathbf{x})}\frac{\partial^A F}{\partial \mathfrak{f}}, \qquad (2.15c)$$

with

$$\frac{\mathfrak{d}^A \mathfrak{f}}{\mathfrak{d}q(\mathbf{x})} = \frac{\partial^A f(\mathbf{x})}{\partial q(\mathbf{x})} - \left(\frac{\partial^A f(\mathbf{x})}{\partial q_{,s}(\mathbf{x})}\right)_{,s}. \qquad (2.15d)$$

Similarly, in a representation in which Ψ is an eigenfunctional of the fields $p(\mathbf{x})$, the operators $q(\mathbf{x})$ are

$$q_j(\mathbf{x}) = i\hbar c\{\mathfrak{d}^A/\mathfrak{d}p^j(\mathbf{x})\}. \qquad (2.16)$$

Because of the existence of fermion fields, the classical theory of gravitation requires "modernization" in another regard. It is impossible to describe the covariant derivatives of the spinors or undors[12,13] describing fermion fields in terms of these fields, the metrical tensor $g_{\mu\nu}$, and derivatives only. The meaning of an undor or spinor is most naturally defined with respect to a local set of pseudo-orthogonal coordinate axes $h_{(0)}$, $h_{(1)}$, $h_{(2)}$, and $h_{(3)}$ given by the four four-vectors $h_{(\alpha)}{}^\mu$ satisfying[14]

$$g_{\mu\nu}h_{(\alpha)}{}^\mu h_{(\beta)}{}^\nu \equiv \mathring{g}_{(\alpha)(\beta)} \equiv \delta_{(\alpha)(\beta)}[1-2\delta_{(\alpha)}{}^{(0)}]. \qquad (2.17)$$

The four unit vectors $h_{(\alpha)}$ form the four "legs" of what Weyl in German calls a *Vierbein* (= fourleg).[15] The gravitational field can be expressed in terms of a chosen "fourleg" field by

$$g^{\mu\nu} \equiv \mathring{g}^{(\alpha)(\beta)}\, h_{(\alpha)}{}^\mu\, h_{(\beta)}{}^\nu, \qquad (2.18a)$$

$$\mathring{g}^{(\alpha)(\beta)}\,\mathring{g}_{(\beta)(\gamma)} \equiv \delta_{(\gamma)}{}^{(\alpha)}, \qquad (2.18b)$$

$$g^{\mu\nu}g_{\mu\lambda} \equiv \delta_\lambda{}^\nu. \qquad (2.18c)$$

The world metric $g_{\mu\nu}$ and the local metric $\mathring{g}_{(\alpha)(\beta)}$ then are used for lowering world and local tensor indices, and their inverses for raising them. The sixteen-component fourleg $h_{(\alpha)}{}^\mu$ is used for projecting a local four-vector onto the world coordinate system, and so on. For simplicity we use a so-called *uniform representation*[13] of undors, and similarly for spinors. This means that with respect to the chosen local axes $h_{(\alpha)}$ we choose for the Dirac matrices the conventional representation independent of coordinates x and satisfying the relations

$$\gamma^{(\alpha)} \equiv -i\beta\alpha^{(\alpha)}, \qquad (2.19a)$$

$$\gamma^{(\alpha)}\gamma^{(\beta)} + \gamma^{(\beta)}\gamma^{(\alpha)} \equiv 2\mathring{g}^{(\alpha)(\beta)}, \qquad (2.19b)$$

$$\beta \equiv \beta^\dagger, \qquad (2.19c)$$

$$\alpha^{(\alpha)} \equiv \alpha^{(\alpha)\dagger}, \qquad (2.19d)$$

$$\gamma^{(\alpha)\dagger}\beta \equiv -\beta\gamma^{(\alpha)}. \qquad (2.19e)$$

We can, for instance, choose the Kramers representation which is also convenient for Yang and Lee's two-component neutrino theory, in which the matrices β and $\alpha^{(n)}$ are represented by ρ_1 and by $\rho_3\sigma^{(n)}$; or we may use the Pauli representation with $\beta = \rho_3$ and $\alpha^{(n)} = \rho_1\sigma^{(n)}$. In both cases

$$\beta^2 \equiv 1, \qquad (2.20a)$$

$$\alpha^{(0)} \equiv -\alpha_{(0)} \equiv 1, \qquad (2.20b)$$

$$\alpha^{(n)} \equiv \gamma_5\sigma^{(n)}. \qquad (2.20c)$$

The Dirac matrices γ^μ satisfying

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu \equiv 2g^{\mu\nu} \qquad (2.21)$$

and therefore depending on x are then given by

$$\gamma^\mu(x) \equiv h_{(\alpha)}{}^\mu(x)\,\gamma^{(\alpha)}. \qquad (2.22)$$

Thus, through (2.18), the gravitational field is determined by the 16-component $h_{(\alpha)}{}^\mu$, as are also through (2.22) the 64 components of the matrices γ^μ. This is far simpler than introducing these 64 components as independent variables as proposed by others.[16,17] We also avoid the necessity of investigating the consequences of the trivial invariance of a theory with unspecified spinor representation with respect to a change of representation. Our method allows only such changes of representation as are given by a change of the fourleg field without change of the local Dirac matrices $\gamma^{(\alpha)}$. Other trivial invariances corresponding to

[12] F. J. Belinfante, Physica 6, 849 (1939).
[13] F. J. Belinfante, Physica 7, 305 (1940). This article also summarizes the theory of *Vierbeine* (fourlegs).
[14] We use the $(-+++)$-signature of the metric; x^0 is the "time" coordinate. Greek indices run from 0 to 3, small italic indices from 1 to 3; without brackets they are world tensor indices, and between brackets ( ) they are local tensor indices indicating components along the fourleg axes. Capital indices are undor indices running from 1 to 4.
[15] H. Weyl, Z. Physik 56, 330 (1929).
[16] E. Schrödinger, Sitzber. preuss. Akad. Wiss. Berlin, Physik-math. Kl. 1932, 105.
[17] J. Heller and P. G. Bergmann, Phys. Rev. 84, 665 (1951).

conservation of the number of fermions minus antifermions, or of the number of heavy fermions minus heavy antifermions, etc., have been discussed in the past.[18] Such invariances under multiplication of field variables by imaginary powers of e are only formally similar to the electromagnetic gauge invariance but really are something independent and different.[19]

In the following we consider the fourleg components $h_{(\alpha)}{}^{\mu}$ as independent variables, and the metrical tensor as quantities derived by (2.18). With the help of the fourleg field we then express the covariant derivatives of the electron fields ψ and $\bar{\psi}$ appearing in the Lagrangian. Our treatment differs from that of the DeWitts[6] by leaving the fourleg field initially arbitrary.

3. PRIMARY CONSTRAINTS

We would like to define "field momenta" by[9]

$$p^{\mathrm{i}} = \partial^P L/\partial q_{\mathrm{i},0} \equiv \mathfrak{p}^{\mathrm{i}}(q,q_{,s},q_{,0}), \qquad (3.1\text{a–b})$$

and then to express the Hamiltonian

$$\mathfrak{H} \equiv \int d^3\mathrm{x} H \qquad (3.2\text{a})$$

with

$$H \equiv p^{\mathrm{i}} q_{\mathrm{i},0} - L(q,q_{,s},q_{,0}) \qquad (3.2\text{b})$$

in terms of the p, q, and $q_{,s}$ only. For this purpose one would want to solve for the $q_{,0}$ in terms of the p from (3.1). However, this is impossible if the Lagrangian $\int L\, d^4x$ is invariant (but for possible boundary effects at the limits of integration) under groups of transformations with "descriptors" ξ which are functions of position.[20–24]

The theory of electrons interacting with gravitational and electromagnetic fields may be derived from the basic Lagrangian

$$L_{\circ} \equiv L_g + L_e + L_i + L_f, \qquad (3.3\text{a})$$

$$L_g \equiv (c^4/16\pi G) j g^{\alpha\rho}(\{\rho\sigma,\beta\}\{\alpha\beta,\sigma\} - \{\alpha\rho,\beta\}\{\beta\sigma,\sigma\}), \qquad (3.3\text{b})$$

$$L_e \equiv -\hbar c j[\kappa\bar{\psi}\psi + \tfrac{1}{2}\bar{\psi}\gamma^{(\alpha)} h_{(\alpha)}{}^{\mu}\nabla_\mu\psi - \tfrac{1}{2}(\nabla_\mu\bar{\psi})h_{(\alpha)}{}^{\mu}\gamma^{(\alpha)}\psi], \qquad (3.3\text{c})$$

$$L_i \equiv \mathfrak{j}^{\mu} A_\mu, \qquad (3.3\text{d})$$

$$L_f \equiv -(16\pi)^{-1} j g^{\alpha\mu} g^{\beta\nu} F_{\alpha\beta} F_{\mu\nu}, \qquad (3.3\text{e})$$

where

$$j \equiv [-g]^{\frac{1}{2}}, \qquad (3.4\text{a})$$

$$\{\rho\sigma,\beta\} \equiv g^{\beta\lambda}[\rho\sigma,\lambda], \qquad (3.4\text{b})$$

$$[\rho\sigma,\lambda] \equiv \tfrac{1}{2}(g_{\lambda\rho,\sigma} + g_{\lambda\sigma,\rho} - g_{\rho\sigma,\lambda}), \qquad (3.4\text{c})$$

$$q_{,\mu} \equiv \partial q/\partial x^\mu, \qquad (3.4\text{d})$$

$$\bar{\psi} \equiv \psi^\dagger \beta, \qquad (3.4\text{e})$$

$$\nabla_\mu\psi \equiv \psi_{,\mu} + \tfrac{1}{4}h_{(\alpha)\lambda}[h_{(\beta)}{}^{\lambda}{}_{,\mu} + h_{(\beta)}{}^{\nu}\{\nu\mu,\lambda\}] \times\gamma^{(\alpha)}\gamma^{(\beta)}\psi, \qquad (3.4\text{f})$$

$$\nabla_\mu\bar{\psi} \equiv \bar{\psi}_{,\mu} + \tfrac{1}{4}h_{(\alpha)\lambda}[h_{(\beta)}{}^{\lambda}{}_{,\mu} + h_{(\beta)}{}^{\nu}\{\nu\mu,\lambda\}] \times\bar{\psi}\gamma^{(\beta)}\gamma^{(\alpha)}, \qquad (3.4\text{g})$$

$$\mathfrak{j}^\mu \equiv ie\, j\, h_{(\alpha)}{}^{\mu}\,\bar{\psi}\gamma^{(\alpha)}\psi \equiv e\, j\, h_{(\alpha)}{}^{\mu}\,\psi^\dagger\,\alpha^{(\alpha)}\,\psi, \qquad (3.4\text{h})$$

$$F_{\mu\nu} \equiv \nabla_\mu A_\nu - \nabla_\nu A_\mu \equiv A_{\nu,\mu} - A_{\mu,\nu}. \qquad (3.4\text{i})$$

Thence,

$$L_g \equiv (c^4/64\pi G)\, j\, (2g^{\lambda\rho}g^{\mu\tau}g^{\nu\sigma} - 2g^{\lambda\mu}g^{\nu\rho}g^{\sigma\tau} - g^{\lambda\sigma}g^{\mu\rho}g^{\nu\tau} + g^{\lambda\mu}g^{\rho\sigma}g^{\nu\tau})g_{\lambda\mu,\nu}g_{\rho\sigma,\tau}, \qquad (3.5\text{a})$$

$$L_e \equiv -\hbar c\, j\,[\kappa\bar{\psi}\psi + \tfrac{1}{2}\bar{\psi}\gamma^\mu\psi_{,\mu} - \tfrac{1}{2}\bar{\psi}_{,\mu}\gamma^\mu\psi + \tfrac{1}{8}\bar{\psi}(\gamma^\mu\gamma_\lambda\gamma^{(\alpha)} - \gamma^{(\alpha)}\gamma_\lambda\gamma^\mu)\psi\, h_{(\alpha)}{}^{\lambda}{}_{,\mu}]. \qquad (3.5\text{b})$$

Our basic Lagrangian $L_{\circ}$ suffers from invariance under gauge transformations

$$A'_\mu \equiv A_\mu + \xi_{,\mu}, \qquad (3.6\text{a})$$

$$\psi' \equiv [\exp(\xi ie/\hbar c)]\psi, \qquad (3.6\text{b})$$

$$\bar{\psi}' \equiv \bar{\psi}\exp(\xi e/i\hbar c), \qquad (3.6\text{c})$$

which for infinitesimal values of the arbitrary "descriptor" function $\xi(x)$ take the form of the transformations discussed by Bergmann and collaborators[22,25] as

$$\delta A_\mu \equiv \xi_{,\mu}, \qquad (3.7\text{a})$$

$$\delta\psi \equiv (ie/\hbar c)\psi\xi, \qquad (3.7\text{b})$$

$$\delta\bar{\psi} \equiv (e/i\hbar c)\bar{\psi}\xi. \qquad (3.7\text{c})$$

Thence the existence follows of the *gauge constraint*[26,27]

$$\phi \equiv P^0 = 0, \qquad (3.8)$$

if P^μ are the momenta conjugated to A_μ.

Our Lagrangian also suffers from invariance under fourleg transformations. Under infinitesimal rotations of the fourlegs through "angles" in the four-dimensional local space at x given by the infinitesimal descriptors $\xi_{(\alpha)(\beta)}(x) = -\xi_{(\beta)(\alpha)}(x)$, the various field variables

[18] See Sec. 2 of F. J. Belinfante, Physica 7, 449 (1940).
[19] F. J. Belinfante, reference 13, Appendix.
[20] L. Rosenfeld, Ann. Physik 5, 113 (1930); Ann. inst. Henri Poincaré 2, 25 (1932).
[21] P. G. Bergmann, Phys. Rev. 75, 680 (1949).
[22] J. L. Anderson and P. G. Bergmann, Phys. Rev. 83, 1018 (1951).
[23] R. H. Penfield, Phys. Rev. 84, 737 (1951).
[24] Pirani, Schild, and Skinner, Phys. Rev. 87, 452 (1952).

[25] P. G. Bergmann and R. Schiller, Phys. Rev. 89, 4 (1953).
[26] The left-hand members of the primary constraints indicated by ϕ (where necessary with indices) are given as functions of the p and q; these functions were easily determined by Bergmann's method. The right-hand members of these constraints are zero. If the p in the left-hand members were interpreted by (3.1) with L replaced by our basic Lagrangian $L_{\circ}$, the constraints would be identities $0\equiv 0$. In the next section we introduce a different interpretation of our field momenta p; the constraints then become "weak equations," and therefore we write = instead of ≡.
[27] "Constraints" are relations which hold automatically between the q and p on a surface time = constant. "Gauge" constraint does not mean constraint on the gauge, but constraint resulting from gauge invariance. Similarly for "fourleg" or "coordinate" constraints.

undergo infinitesimal transformations

$$\delta h_{(\alpha)}{}^{\mu} \equiv \xi_{(\alpha)(\beta)} h^{(\beta)\mu}, \quad (3.9a)$$

$$\delta A_{\mu} \equiv 0, \quad (3.9b)$$

$$\delta\psi \equiv \tfrac{1}{4}\xi_{(\alpha)(\beta)}\gamma^{(\alpha)}\gamma^{(\beta)}\psi, \quad (3.9c)$$

$$\delta\bar{\psi} \equiv \tfrac{1}{4}\bar{\psi}\gamma^{(\beta)}\gamma^{(\alpha)}\xi_{(\alpha)(\beta)}, \quad (3.9d)$$

again of the general form discussed by Bergmann. Thence the existence follows of the six *fourleg constraints*[27,28]

$$\phi^{(\alpha)(\beta)} \equiv [\tfrac{1}{4}\bar{\pi}(\gamma^{(\alpha)}\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^{(\alpha)})\psi + \tfrac{1}{4}\bar{\psi}(\gamma^{(\alpha)}\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^{(\alpha)})\pi + p_{\mu}{}^{(\alpha)} h^{(\beta)\mu} - p_{\mu}{}^{(\beta)} h^{(\alpha)\mu}] = 0. \quad (3.10)$$

Here, $\bar{\pi}$ is the row matrix of momenta conjugate to the components of the column matrix ψ, and the column matrix π is the momentum conjugate to the row matrix $\bar{\psi}$. The $p_{\mu}{}^{(\alpha)}$ are the momenta conjugate to the fourleg components $h_{(\alpha)}{}^{\mu}$.

Finally, $\int L_o d^4x$ suffers from invariance (but for boundary effects) under arbitrary coordinate transformations $x'^{\mu} = x^{\mu} + \xi^{\mu}(x)$, under which, for infinitesimal values of the descriptors ξ^{μ}, the field variables locally undergo transformations

$$\delta h_{(\alpha)}{}^{\mu} \equiv \xi^{\mu}{}_{,\nu} h_{(\alpha)}{}^{\nu}, \quad (3.11a)$$

$$\delta A_{\nu} \equiv -\xi^{\mu}{}_{,\nu} A_{\mu}, \quad (3.11b)$$

$$\delta\psi \equiv 0, \quad (3.11c)$$

or, if one considers the alteration of the field at given coordinate values rather than locally at a given point

$$\bar{\delta} h_{(\alpha)}{}^{\mu} \equiv \xi^{\mu}{}_{,\nu} h_{(\alpha)}{}^{\nu} - \xi^{\nu} h_{(\alpha)}{}^{\mu}{}_{,\nu}, \quad (3.12a)$$

$$\bar{\delta} A_{\nu} \equiv -\xi^{\mu}{}_{,\nu} A_{\mu} - \xi^{\mu} A_{\nu,\mu}, \quad (3.12b)$$

$$\bar{\delta}\psi \equiv -\xi^{\nu}\psi_{,\nu}, \quad (3.12c)$$

$$\bar{\delta}\bar{\psi} \equiv -\xi^{\nu}\bar{\psi}_{,\nu}. \quad (3.12d)$$

From this invariance property of the Lagrangian, the existence follows of the four *coordinate constraints*[27]

$$0 = \phi_{\mu} \equiv p_{\mu}{}^{(\alpha)} h_{(\alpha)}{}^{0} - P^{0} A_{\mu} + (c^4/32\pi G)\, j\, g_{\alpha\beta,s} \times [2g^{\alpha 0} g^{\beta 0}\delta_{\mu}{}^{s} - 2g^{\alpha 0} g^{\beta s}\delta_{\mu}{}^{0} - g^{\alpha\beta} g^{00}\delta_{\mu}{}^{s} + g^{\alpha\beta} g^{0s}\delta_{\mu}{}^{0}]$$
$$\equiv p_{\mu}{}^{(\alpha)} h_{(\alpha)}{}^{0} - \phi A_{\mu} + (c^4/16\pi G)[j(g^{0s}\delta_{\mu}{}^{0} - g^{00}\delta_{\mu}{}^{s})]_{,s}. \quad (3.13)$$

The constraints (3.8), (3.10), and (3.13) are the *primary constraints* of our theory. Other constraints (i.e., relations between the p and the q) will be derived in following sections.

[28] A. Schild and R. Skinner, Carnegie Institute of Technology, Department of Mathematics, Technical Report No. 23, September 10, 1952, using in this regard a notation similar to that used by the DeWitts,[6] write Π and $-\bar{\Pi}$ for our $\bar{\pi}$ and π. The minus sign is due to their lack of distinction between antederivatives and postderivatives. Notice that in our notation $\bar{\pi} = -\pi^{\dagger}\beta$, as we defined all momenta as postderivatives (3.1a) for having $pq_{,0}$ and not $q_{,0}p$ in (3.2b).

4. WEAK EQUATIONS

If we would define momenta p^{i} by

$$p^{i} \equiv \partial^{P} L_o / \partial q_{i,0}, \quad (4.1)$$

the constraints would be identities

$$\phi^{k}(p, q, q_{,s}) \equiv \phi^{k}((\partial^{P} L_o/\partial q_{,0}), q, q_{,s}) \equiv 0, \quad (4.2a\text{–}b)$$

where (4.2a) is due to (4.1) and (4.2b) is due to the invariance properties of L_o. Some effects of identities such as between the first and last member of (4.2), on the canonical formulation of the theory, would be (i) the q and p cannot arbitrarily be prescribed at some initial time x^0 and (ii) the relations (4.2) between the p^{i} defined by (4.1) show that (4.1) are not independent, and therefore the $q_{,0}$ are not determined uniquely by (4.1). Therefore, for an *allowable* choice of the q and p [satisfying (4.2)] there still is arbitrariness in the $q_{,0}$.

Worst of all, identities of the form of (4.2) would not be compatible with the canonical commutation relations (2.1). This truly should mean that these relations (2.1) need alteration; but, as it is hard to guess how this should be done, we postpone this question to Part III, Sec. 24. Instead, we first try to maintain a quantization of variables q and p as closely similar to (2.1) as possible, avoiding the discrepancy with (4.2) by assuming with Dirac that the p in our commutation relations differ from (4.1) by unknown q numbers η^{i} which in some ill-understood way "have the value zero" in all physical states of nature:

$$p^{i} \equiv \partial^{P} L_o/\partial q_{i,0} + \eta^{i}, \quad (4.3a)$$

$$\eta^{i} = 0. \quad (4.3b)$$

The second equation here is called a "weak" equation by Dirac.[5] Dirac assumes that also the Lagrangian field equations (or at least some of them) may be valid only weakly[29]:

$$\partial^{P} L_o/\partial q_{i} - (\partial^{P} L_o/\partial q_{i,\mu})_{,\mu} \equiv \zeta^{i}, \quad \zeta^{i} = 0. \quad (4.4a\text{–}b)$$

The constraints, being identities between the q, the $q_{,s}$, and the $(\partial^{P} L_o/\partial q_{,0})$, then take the form

$$\phi^{k}(p-\eta, q, q_{,s}) \equiv 0. \quad (4.5)$$

Since $\phi^{k}(p, q, q_{,s}) - \phi^{k}(p-\eta, q, q_{,s})$ is a polynomial containing at least one factor η in every term, this dif-

[29] Sometimes, indeed, even if momenta are defined by (4.3) instead of (4.1), some of the field equations may lead to constraints. If, for instance, one uses for the Maxwell field a first-order Lagrangian $L_f \equiv (4\pi)^{-1} j\, g^{\alpha\mu} g^{\beta\nu} F_{\alpha\beta}[\tfrac{1}{4}F_{\mu\nu} + A_{\mu,\nu}]$, in which the $F_{\mu\nu} \equiv -F_{\nu\mu}$ and the A_{μ} are considered to be independent variables q, the field equations for F_{rs} would be $F_{rs} + A_{r,s} - A_{s,r} = 0$ and would endanger the commutativity postulated by (2.1) in such a theory between F_{rs} and the canonical conjugate to A_r, unless either we replace 0 in the field equation by ζ as in (4.4), or we change (2.1). Dirac first introduces the ζ, and then, by a formula expressed in terms of the ζ, alters (2.1) in such a way that ζ can again be omitted wherever this is possible. This procedure leads to results identical with those obtained by Belinfante's method of quantization of a theory in which there are so-called "derived variables"; see reference 10, and compare our Sec. 12.

ference will vanish weakly by (4.3b) *if* we assume that both $Q\eta=0$ and $\eta Q=0$ follow whenever $\eta=0$. (4.6)

Then, (4.5) leads to

$$\phi^{\mathrm{k}}(p,q,q_{,s}) \equiv \phi^{\mathrm{k}}(p,\cdots) - \phi^{\mathrm{k}}(p-\eta,\cdots) = 0. \quad \text{(4.7a–b)}$$

Thus the constraints, when expressed in terms of the new p given by (4.3), hold only weakly,[30] and only as a consequence of (4.3b). The identities (4.7a) express the constraint functions ϕ^{k} in terms of the η.

Dirac's theory is vague about the meaning of the η and the ζ. This is an advantage in as far as it shows the independence of many results, of whatever the η and ζ may be. It is a handicap when the covariance of the theory is to be proved. [See Sec. 13.]

Worse is Dirac's vagueness about what is meant by "weakly zero." If this means a condition imposed upon the state vector, it is not clear why the condition (4.6) would be fulfilled. It may therefore be preferable to understand weak equations as relations singling out a subspace of Hilbert space with the idea that whatever the theory would predict outside this physical subspace of Hilbert space should not be believed.[31] In this case, any Q which does not commute with some η (or ζ) would then be a physically meaningless quantity, even if expressible in terms of our original q and p. How one might be able to formulate such a theory in more detail is discussed in Sec. 24.

5. SECOND-CLASS CONSTRAINTS

In Sec. 3 we discussed constraints which were due to invariance properties of the Lagrangian. A quite different type of constraints exists on account of the fact that the Lagrangian is linear in the time-derivatives of ψ and $\bar{\psi}$, with coefficients not depending on any other $q_{,0}$. [Compare (3.5b).] Therefore, the corresponding definitions (4.3a) give the undors

$$\bar{\eta} \equiv \bar{\pi} - \partial^P L_\circ/\partial\psi_{,0} \equiv \bar{\pi} + \tfrac{1}{2}\hbar c\, j\, \bar{\psi}\gamma^0, \quad \text{(5.1a)}$$

$$\eta \equiv \pi - \partial^P L_\circ/\partial\bar{\psi}_{,0} \equiv \pi + \tfrac{1}{2}\hbar c\, j\, \gamma^0\psi \quad \text{(5.1b)}$$

as functions of the q and p. This makes constraints out of the equations (4.3b). Since these constraints have nonvanishing (modernized) P.B. with each other and with some of the other constraints, we call them *second-class constraints* in Dirac's terminology,[5] and indicate these by θ_n. So, $\theta_1, \theta_2, \theta_3, \theta_4$ are the components of η, and $\theta_5, \theta_6, \theta_7, \theta_8$ are the components of $\bar{\eta}$.

6. HAMILTONIAN

By hard work it can now be shown that the Hamiltonian (3.2) with L replaced by $L_\circ$ is equal to

$$\mathcal{H}_\circ \equiv \int d^3\mathrm{x}\, H_\circ(p,q,q_{,s},q_{,0}) \quad \text{(6.1a)}$$

with the integrand given by

$$H_\circ \equiv \mathfrak{h} + \phi^{\mathrm{k}}\beta_{\mathrm{k}} + \theta_{\mathrm{n}}\mu^{\mathrm{n}} + O_2. \quad \text{(6.1b)}$$

Here, $\mathfrak{h}$ is a function of the p, q, and $q_{,s}$ only, given by

$$\begin{aligned}
\mathfrak{h} \equiv\ & (2\pi G/c^4)(g^{00}j)^{-1} p_\mu{}^{(\alpha)} p_\nu{}^{(\beta)}[h_{(\alpha)}{}^\mu h_{(\beta)}{}^\nu \\
& \qquad - h_{(\beta)}{}^\mu h_{(\alpha)}{}^\nu - \dot{g}_{(\alpha)(\beta)}g^{\mu\nu}] \\
& + (2g^{00})^{-1} p_\mu{}^{(\alpha)} g_{\rho\sigma,s}[h_{(\alpha)}{}^\rho g^{s0}g^{\sigma\mu} + h_{(\alpha)}{}^\rho g^{s\sigma}g^{\mu 0} \\
& \qquad - h_{(\alpha)}{}^\rho g^{s\mu}g^{\sigma 0} + h_{(\alpha)}{}^0 g^{s\sigma}g^{\rho\mu} - h_{(\alpha)}{}^s g^{\mu\rho}g^{\sigma 0}] \\
& + (c^4/32\pi G)(g^{00})^{-1} j\, g_{\mu\nu,r}\, g_{\rho\sigma,s}\, [\tfrac{1}{4}g^{\mu\nu}g^{\rho\sigma}g^{r0}g^{s0} \\
& \qquad - \tfrac{1}{4}g^{\mu\nu}g^{\rho\sigma}g^{rs}g^{00} + \tfrac{1}{2}g^{\mu\rho}g^{\nu\sigma}g^{rs}g^{00} - \tfrac{1}{2}g^{\mu\rho}g^{\nu\sigma}g^{r0}g^{s0} \\
& \qquad + 2g^{\mu 0}g^{\nu 0}g^{\rho r}g^{\sigma s} + 2g^{\mu 0}g^{\nu\rho}g^{\sigma r}g^{s0} - g^{\mu 0}g^{\nu\rho}g^{\sigma 0}g^{rs} \\
& \qquad - g^{\mu r}g^{\nu 0}g^{\rho 0}g^{\sigma s} - g^{\mu 0}g^{\nu s}g^{\rho r}g^{\sigma 0} - g^{\mu s}g^{\nu\rho}g^{\sigma r}g^{00}] \\
& - 2\pi(g^{00}j)^{-1} g_{\alpha\beta}P^\alpha P^\beta \\
& + (g^{00})^{-1} A_{\alpha,r}[P^r g^{\alpha 0} - P^\alpha g^{r0} - P^0 g^{\alpha r}] \\
& + (8\pi g^{00})^{-1} j\, A_{\alpha,r} A_{\beta,s}[g^{\alpha\beta}g^{rs}g^{00} - g^{\alpha s}g^{\beta r}g^{00} \\
& \qquad + 2g^{\alpha s}g^{\beta 0}g^{r0} - g^{\alpha\beta}g^{r0}g^{s0} - g^{\alpha 0}g^{\beta 0}g^{rs}] \\
& + \hbar c\, j\, [\kappa\bar{\psi}\psi + \tfrac{1}{2}\bar{\psi}\gamma^r\psi_{,r} - \tfrac{1}{2}\bar{\psi}_{,r}\gamma^r\psi \\
& \qquad + \tfrac{1}{8}\bar{\psi}(\gamma^r\gamma_\mu\gamma^{(\alpha)} - \gamma^{(\alpha)}\gamma_\mu\gamma^r)\psi h_{(\alpha)}{}^\mu{}_{,r}] \\
& - ie\, j\, \bar{\psi}\gamma^\mu\psi A_\mu.
\end{aligned} \quad \text{(6.2)}$$

The ϕ^{k} are given by (3.8), (3.10), and (3.13). Their coefficients β_{k} are given by

$$\beta \equiv (\mathfrak{g}^{00})^{-1}[A_\alpha \mathfrak{g}^{\alpha\beta}]_{,\beta}, \quad \text{(6.3a)}$$

$$\beta_{(\alpha)(\beta)} \equiv \tfrac{1}{2}h_{(\alpha)\mu}h_{(\beta)}{}^0(\mathfrak{g}^{00})^{-1}\mathfrak{g}^{\mu\nu}{}_{,\nu} + \tfrac{1}{4}[h_{(\beta)\mu}h_{(\alpha)}{}^\mu{}_{,0} - h_{(\alpha)\mu}h_{(\beta)}{}^\mu{}_{,0}], \quad \text{(6.3b)}$$

$$\beta^\mu \equiv (\mathfrak{g}^{00})^{-1}\mathfrak{g}^{\mu\beta}{}_{,\beta}, \quad \text{(6.3c)}$$

where $\mathfrak{g}^{\alpha\beta} \equiv j\, g^{\alpha\beta}$. The θ_{n} are given by the column-matrix and the row-matrix undors (5.1b) and (5.1a):

$$\theta \equiv \eta, \quad \text{(6.4a)}$$

$$\bar{\theta} \equiv \bar{\eta} \equiv -\theta^\dagger\beta, \quad \text{(6.4b)}$$

and their coefficients μ^{n} by

$$\bar{\mu} \equiv +\mu^\dagger\beta \equiv \bar{\psi}_{,0} + \tfrac{1}{8}\bar{\psi}(\gamma^{(\alpha)}\gamma_\mu - \gamma_\mu\gamma^{(\alpha)}) \times[h_{(\alpha)}{}^\mu{}_{,0} - h_{(\alpha)}{}^0(\mathfrak{g}^{00})^{-1}\mathfrak{g}^{\mu\beta}{}_{,\beta}], \quad \text{(6.5a–b)}$$

$$\mu \equiv \psi_{,0} + \tfrac{1}{8}[h_{(\alpha)}{}^\mu{}_{,0} - h_{(\alpha)}{}^0(\mathfrak{g}^{00})^{-1}\mathfrak{g}^{\mu\beta}{}_{,\beta}] \times(\gamma_\mu\gamma^{(\alpha)} - \gamma^{(\alpha)}\gamma_\mu)\psi, \quad \text{(6.5c)}$$

so that

$$\theta_{\mathrm{n}}\mu^{\mathrm{n}} \equiv \theta^{\mathrm{T}}\bar{\mu}^{\mathrm{T}} + \bar{\theta}\mu \equiv \bar{\theta}\mu - \bar{\mu}\theta \quad \text{(6.6)}$$

since these undors are F.D. quantities. ($^{\mathrm{T}}$ means transposed.) Then, O_2 is given by

$$O_2 \equiv (2\pi G/c^4)(\mathfrak{g}^{00})^{-1}\eta_\mu{}^{(\alpha)}\eta_\nu{}^{(\beta)}[\dot{g}_{(\alpha)(\beta)}g^{\mu\nu} + h_{(\beta)}{}^\mu h_{(\alpha)}{}^\nu - h_{(\alpha)}{}^\mu h_{(\beta)}{}^\nu] + 2\pi(\mathfrak{g}^{00})^{-1}g_{\alpha\beta}\eta^\alpha\eta^\beta, \quad \text{(6.7)}$$

where $\eta_\mu{}^{(\alpha)}$ and η^α are given by (4.3a) with $q_{\mathrm{i}} = h_{(\alpha)}{}^\mu$ and with $q_{\mathrm{i}} = A_\alpha$.

While $H_\circ$ is determined by (3.2) but for a three-dimensional divergence term, the subdivision of $H_\circ$ according to (6.1b) into a canonical term, constraint terms, and a term of second order in weakly vanishing quantities is not unique. We may add to β_{k} or to μ^{n}

---

[30] Compare reference 26.

[31] F. J. Belinfante, Physica 12, 17 (1946).

any functions of the q, $q_{,s}$, and p, and then compensate the resulting terms by subtracting them from $\mathfrak{h}$. Or, we may add to β_k or to μ^n any weakly vanishing quantities, and then compensate the resulting terms in $H_\circ$ by subtracting these terms from O_2. Moreover, instead of expressing $H_\circ$ in terms of the constraints ϕ^k and θ_n, we could make use of some linear combinations of them. In particular, we can achieve some simplifications in the following by replacing the ϕ^k by constraints Φ^k defined by

$$\Phi \equiv \phi \equiv P^0, \tag{6.8a}$$

$$\begin{aligned}\Phi^{(\alpha)(\beta)} &\equiv \phi^{(\alpha)(\beta)} + \tfrac{1}{4}\bar{\theta}(\gamma^{(\beta)}\gamma^{(\alpha)} - \gamma^{(\alpha)}\gamma^{(\beta)})\psi \\ &\quad + \tfrac{1}{4}\bar{\psi}(\gamma^{(\beta)}\gamma^{(\alpha)} - \gamma^{(\alpha)}\gamma^{(\beta)})\theta \\ &\equiv p_\mu{}^{(\alpha)}h^{(\beta)\mu} - p_\mu{}^{(\beta)}h^{(\alpha)\mu} \\ &\quad + \tfrac{1}{4}\hbar c\, j\, \bar{\psi}[\gamma^{(\alpha)}\gamma^0\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^0\gamma^{(\alpha)}]\psi,\end{aligned} \tag{6.8b}$$

$$\begin{aligned}\Phi_\mu &\equiv \phi_\mu + \phi A_\mu \\ &\equiv p_\mu{}^{(\alpha)}h_{(\alpha)}{}^0 + (c^4/16\pi G) \\ &\qquad \times[\mathfrak{g}^{0s}\delta_\mu{}^0 - \mathfrak{g}^{00}\delta_\mu{}^s]_{,s}.\end{aligned} \tag{6.8c}$$

The Hamiltonian can then be written as

$$\mathfrak{H}_\circ \equiv \int d^3\mathbf{x}\{\mathfrak{H} + \Phi^k B_k + \theta_n M^n + O_2\}; \tag{6.9}$$

$$\Phi^k B_k \equiv \Phi B + \Phi_\mu B^\mu + \Phi^{(\alpha)(\beta)}B_{(\alpha)(\beta)}, \tag{6.10a}$$

$$B \equiv (g^{00})^{-1}\, g^{\alpha 0}A_{\alpha,0}, \tag{6.10b}$$

$$B^\mu \equiv (\mathfrak{g}^{00})^{-1}\, \mathfrak{g}^{\mu\beta}{}_{,\beta} \equiv \beta^\mu, \tag{6.10c}$$

$$B_{(\alpha)(\beta)} \equiv \beta_{(\alpha)(\beta)} \quad \text{[see Eq. (6.3b)]}; \tag{6.10d}$$

$$\theta_n M^n \equiv \bar{\theta}M - \bar{M}\theta, \quad \text{[compare Eq. (6.6)]}, \tag{6.11a}$$

$$M \equiv \psi_{,0}, \tag{6.11b}$$

$$\bar{M} \equiv \bar{\psi}_{,0}; \tag{6.11c}$$

$$\mathfrak{H} \equiv \mathfrak{h} + (g^{00})^{-1}\, P^0 g^{\alpha r}A_{\alpha,r}, \tag{6.12}$$

with $\mathfrak{h}$ from (6.2).

The formulas given so far for the Hamiltonian are valid in classical approximation only; that is, we have not really considered here the question of the correct ordering of the factors in each term, except for a distinction between even and odd permutations of Fermi-Dirac quantities necessary in a modernized theory in connection with the sign of the term. We did not even introduce the "symmetrization" conventional in field theory[10] according to which for instance a product $p^i Q_i(q)$ should be written as

$$\tfrac{1}{2}\{p^i Q_i + (-)^{u_i}\, Q_i p^i\}. \tag{6.13}$$

One reason for this negligence is that on account of experience with the quantization of the "linear" theory of gravitation[7–8] we are not even convinced that there *exists* a correct ordering of factors in the Hamiltonian as long as we try to express it in terms of the variables q and p used so far. We believe that one should look for the correct ordering of factors *after* the "canonization" of the variables discussed in Sec. 12.

7. HAMILTON'S EQUATIONS

Regarding the Hamiltonian density $H_\circ$ as a function of *independent* variables q, $q_{,s}$, $q_{,0}$, and p, we find [compare (3.2) and (2.15d)]

$$q_{i,0} \equiv \partial^A H_\circ/\partial p^i \equiv \mathfrak{d}^A \mathfrak{H}_\circ/\mathfrak{d}p^i. \tag{7.1}$$

The Lagrangian field equations (4.4a) by (4.3a), (3.2), and (2.15d) can be written in the form

$$p^i{}_{,0} \equiv o^i - \mathfrak{d}^P \mathfrak{H}_\circ/\mathfrak{d}q_i, \tag{7.2}$$

where

$$o^i \equiv \eta^i{}_{,0} - \zeta^i \quad (=0 \text{ weakly}). \tag{7.3a–b}$$

Now, let F be a function $F(q, q_{,s}, p, p_{,r})$ not explicitly depending on the field velocities $q_{,0}$ and $p_{,0}$. Then

$$\begin{aligned}F_{,0} \equiv (\partial^P F/\partial q_i)q_{i,0} &+ (\partial^P F/\partial q_{i,s})q_{i,0s} \\ &+ p^i{}_{,0}(\partial^A F/\partial p^i) + p^i{}_{,0r}(\partial^A F/\partial p^i{}_{,r}),\end{aligned} \tag{7.4}$$

if we understand products like $(\partial^P F/\partial q)q_{,0}$ as meaning that in the terms of F we place the factor $q_{,0}$ at the position where we differentiate away the factor q, etc. Now insert for $q_{i,0}$ and for $p^i{}_{,0}$ the expressions (7.1)–(7.2), and for $q_{i,0s}$ and for $p^i{}_{,0r}$ their spatial derivatives. The resulting expression we compare with the integral over $\mathbf{x}'$ of (2.10) with A replaced by F and with B replaced by $H_\circ$, keeping (2.15) in mind. Thus we find

$$F_{,0} \equiv (F, \mathfrak{H}_\circ) + o^i(\partial^A F/\partial p^i) + o^i{}_{,r}(\partial^A F/\partial p^i{}_{,r}), \tag{7.5}$$

where again the last terms should be understood as substitutions of o^i at the positions where factors p^i were differentiated away. On account of (7.3b), the equations (7.5) give *weakly* Hamilton's equations of motion.

In (7.1)–(7.2) we differentiated $H_\circ$ at constant $q_{i,0}$. From (3.2) and (4.3a) it is easily seen that

$$\mathfrak{d}^P \mathfrak{H}_\circ/\mathfrak{d}q_{i,0} \equiv \partial^P H_\circ/\partial q_{i,0} \equiv \eta^i. \tag{7.6}$$

8. SECONDARY CONSTRAINTS

We now apply (7.5) to the constraint functions Φ^k and θ_n. As the constraints according to (4.7b) and (6.4) with (4.3b) are weak equations valid for any time x^0, also their time derivatives must be valid as weak equations. Combining this with the weak vanishing of o^i by (7.3), we obtain the weak equations

$$(\Phi^k, \mathfrak{H}_\circ) = 0 \text{ (weakly)}; \tag{8.1a}$$

$$(\theta_n, \mathfrak{H}_\circ) = 0. \tag{8.1b}$$

In Secs. 9 to 11 we discuss methods of evaluating $(\Phi^k, \mathfrak{H}_\circ)$. We find there that (8.1a) is equivalent to a set of weak equations

$$\chi^k = 0, \tag{8.2}$$

where the χ^k are again functions of the q, $q_{,s}$, and p. The equations (8.2) are known as the *secondary constraints*.[6]

One finds for χ, χ_μ, and $\chi^{(\alpha)(\beta)}$ [32]

$$\chi \equiv P^s{}_{,s} + ie\, j\, \bar{\psi}\gamma^0\psi; \tag{8.3a}$$

$$\chi_0 \equiv \mathfrak{H} + [p_0{}^{(\alpha)}\, h_{(\alpha)}{}^s - P^s A_0]_{,s} + C\mathfrak{g}^{rs}{}_{,rs}, \tag{8.3b}$$

$$\chi_s \equiv -\mathfrak{G}_s + [p_s{}^{(\alpha)}\, h_{(\alpha)}{}^r - P^r A_s]_{,r} - C\mathfrak{g}^{0r}{}_{,rs}; \tag{8.3c}$$

$$\chi^{(\alpha)(\beta)} \equiv 0. \tag{8.3d}$$

Here, we put

$$C \equiv c^4/16\pi G, \tag{8.4}$$

$$-\mathfrak{G}_s \equiv p_\lambda{}^{(\alpha)}\, h_{(\alpha)}{}^\lambda{}_{,s} + P^\lambda A_{\lambda,s} + \tfrac{1}{2}\hbar c\, j(\bar{\psi}_{,s}\gamma^0\psi - \bar{\psi}\gamma^0\psi_{,s}). \tag{8.5}$$

Equations (8.3) show that there are only five secondary constraints corresponding to the eleven primary constraints $\Phi^k = 0$ and obtained by postulating the validity of the latter for all time. If we now repeat this procedure and use the fact that also the secondary constraints (8.2) must be valid for all time, we find the weak equations

$$0 = \chi^k{}_{,0} = (\chi^k, \mathfrak{H}_o). \tag{8.6}$$

We find in Sec. 11 that these equations are satisfied automatically and do not give rise to further constraints.[32]

Returning now to (8.1b), let us remember how the θ_n were introduced. Let q^n be those field variables q_i for which

$$\partial^P L_o/\partial q^n{}_{,0} \equiv f_n(q) \tag{8.7}$$

does not depend on any $q_{,\mu}$. (So, ψ and $\bar{\psi}$ are the q^n.) Let us denote the corresponding p^i, η^i, ζ^i, o^i by p_n, θ_n, ζ_n, and o_n. Then the (4.3) become constraints

$$p_n - \partial^P L_o/\partial q^n{}_{,0} \equiv \theta_n(q,p) = 0. \tag{8.8a–b}$$

By (7.5) and (8.8) we then find

$$(\theta_n, \mathfrak{H}_o) \equiv \theta_{n,0} - o^i(\partial^P\theta_n/\partial p^i) \equiv \theta_{n,0} - o^i\delta_i{}^n \equiv \zeta_n, \tag{8.9a–c}$$

where we used (7.3a).

In our particular theory, it is allowable and covariant to assume that among the Lagrangian field equations the electron wave equations are valid strongly. By (4.4) this means assuming

$$\zeta_n \equiv 0, \tag{8.10a}$$

so that (8.9) yields

$$(\theta_n, \mathfrak{H}_o) \equiv 0. \tag{8.10b}$$

For this reason, the validity of the second-class constraints (8.8) at all times by (8.1b) does not lead to any new secondary constraints.

By (8.2), the three-dimensional volume integrals of (8.3) over space within a closed surface Σ weakly vanish. This yields, by (3.4h), for the charge of the universe the limit for $\Sigma \to \infty$ of the surface integral of $-P^s$ over Σ. Similarly, by (6.9) and the weak vanishing of Φ^k, θ_n, and O_2, we find that if $\mathfrak{g}^{rs}{}_{,r} \to 0$ at spatial infinity, then the total energy in the universe is weakly equal to the limit for $\Sigma \to \infty$ of the surface integral of $[P^s A_0 - p_0{}^{(\alpha)}\, h_{(\alpha)}{}^s]$. As $\mathfrak{G}_s$ is easily recognized as the "canonical" or "orbital" momentum density,[33] (8.3c) may be used similarly for expressing the total momentum of the universe as an integral over a closed surface at spatial infinity.

9. MODIFIED POISSON BRACKETS

From (6.4) with (5.1) it is easily seen that the modernized Poisson brackets (2.10) between the θ_n do not all vanish[14]:

$$(\theta_K(\mathbf{x}), \bar{\theta}^L(\mathbf{x}')) \equiv \hbar c\, j\, (\gamma^0)_K{}^L\, \delta_3(\mathbf{x}-\mathbf{x}'). \tag{9.1}$$

Now, if we would try to quantize our fields according to (2.1), then (2.9) would follow, and then we would find nonvanishing anticommutators

$$[\theta_K(\mathbf{x}); \bar{\theta}^L(\mathbf{x}')] \equiv i\hbar^2 c^2\, j\, (\gamma^0)_K{}^L\, \delta_3(\mathbf{x}-\mathbf{x}') \tag{9.2a}$$

incompatible with the simultaneous weak vanishing of θ_K and of $\bar{\theta}^L$ which requires

$$[\theta_K(\mathbf{x}); \bar{\theta}^L(\mathbf{x}')] = 0 \tag{9.2b}$$

instead. In order to achieve (9.2b) and to prevent (9.2a), we need a modified scheme of quantization.

Such a modification was suggested for Einstein-Bose fields by Dirac[5] in his work on second-class constraints By "modernization" we adapt his method to our Fermi-Dirac field of electrons. We modify the quantization scheme by replacing (2.9) by

$$\lim_{\hbar\to 0} \frac{[A; B']}{i\hbar c} = [A, B'], \tag{9.3}$$

where $[A,B']$, the so-called (modernized) *modified Poisson bracket* of A and B', is now defined in terms of the ordinary (modernized) P.B., by

$$[A,B'] \equiv (A,B') - \tfrac{1}{2}\int d^3\mathbf{x}''\int d^3\mathbf{x}'''(A,\theta_m'') \times c^{mn}(\mathbf{x}'',\mathbf{x}''')(\theta_n''',B') + \tfrac{1}{2}(-)^{u_A u_B} \times d^3\mathbf{x}''\int d^3\mathbf{x}'''(B',\theta_n''')c^{nm}(\mathbf{x}''',\mathbf{x}'')(\theta_m'',A). \tag{9.4}$$

Here, A means $A(\mathbf{x})$, B' means $B(\mathbf{x}')$, etc., all for the same x^0. The matrix c^{mn} in (9.4) is the inverse of the matrix

$$C_{mn}(\mathbf{x}',\mathbf{x}) \equiv (\theta_m(\mathbf{x}'),\theta_n(\mathbf{x})) \equiv -(-)^{u_m u_r} C_{nm}(\mathbf{x},\mathbf{x}'), \tag{9.5}$$

according to

$$\int d^3\mathbf{x}''' c^{lm}(\mathbf{x}'',\mathbf{x}''')C_{mn}(\mathbf{x}''',\mathbf{x}') \equiv \delta_n{}^l\,\delta_3(\mathbf{x}''-\mathbf{x}'). \tag{9.6}$$

In the classical limit $[A\ ; B'] \to 0$, in which

$$c^{ml}C_{nm} = (-)^{(u_l+u_m)(u_m+u_n)}C_{nm}c^{ml}, \tag{9.7}$$

we also have

$$\int d^3\mathbf{x}''' C_{nm}(\mathbf{x}',\mathbf{x}''')c^{ml}(\mathbf{x}''',\mathbf{x}'') = \delta_n{}^l\delta_3(\mathbf{x}'-\mathbf{x}''). \tag{9.8}$$

[32] W. L. Kennedy, Ph.D. thesis, Purdue University (1957).

[33] F. J. Belinfante, Physica 6, 887 (1939); 7, 449 (1940).

Consistency of (9.8) with (9.5) through (9.7) is ensured by

$$c^{lm}(\mathbf{x}'',\mathbf{x}''') = (-)^{(u_l+1)(u_m+1)}\, c^{ml}(\mathbf{x}''',\mathbf{x}''). \quad (9.9)$$

In the classical limit, in which by (2.2) with (2.5) all quantities become purely commutative or anticommutative according to

$$AB' = (-)^{u_A u_B}\, B'A, \quad (9.10)$$

the modified P.B. (9.4) satisfy the same identities (2.11)–(2.13) as the (modernized) regular P.B. (2.10). Since also "commutators" (2.5) satisfy the identities (2.11)–(2.12) and satisfy (2.13) but for terms which in the classical limit vanish even after division by $\hbar$, quantization by (9.3) would seem feasible.

Now, let F be *any* function of the q and p and their spatial derivatives. Then it follows from (9.4) with (9.5)–(9.6), with (2.11) and with (9.8) that, in the classical limit (9.10),

$$[F,\theta_n{}'] = 0. \quad (9.11)$$

Therefore, (9.3) suggests

$$[F\,;\theta_n{}'] \equiv 0, \quad (9.12)$$

which means that θ_n should be commutative or anticommutative with "everything", and therefore should be a quasi-c-number. As we knew already that at least weakly the θ_n should vanish, we conclude that the θ_n should be vanishing quasi-c-numbers and therefore simply would be c-numbers (as the sign of zero is irrelevant) and thus would vanish strongly:

$$\theta_n \equiv 0 \quad [\text{so } \theta \equiv \eta \equiv 0 \quad \text{and} \quad \bar{\theta} \equiv \bar{\eta} \equiv 0]. \quad (9.13)$$

As we put the θ_n strongly equal to zero, we have to be careful *not* to conclude from that that there would be no difference between regular and modified P.B. Although the θ_n vanish strongly, their regular P.B. with the various p^i and q_i (or with each other) do *not* vanish. That is, (8.8a) with (9.13) tell us that the p_n are strongly equal to the $f_n(q)$ of (8.7); yet, in defining the regular P.B., we introduced differentiations with respect to p at constant q, and with respect to q at constant p, obviously ignoring the constraints (9.13). Thus, for instance, the regular P.B. of (8.8) with q_i yields

$$(q_i,\theta_n) = \partial p_n/\partial p^i = \delta_{in}, \quad (9.14a)$$

and, similarly,

$$(\theta_n,p^i) = \partial^P\theta_n/\partial q_i = -\,\partial^P f_n/\partial q_i. \quad (9.14b)$$

From (9.13)–(9.14) it is clear that (modernized) regular P.B. of any function F of the original q and p can no longer be calculated as soon as the function F is simplified by any use of the constraints (9.13). On the other hand, because of (9.11), there are no difficulties with the modified P.B. Therefore, it will be our procedure, before any use of (9.13) is made, to calculate all the modified P.B. by (9.4) and, once this has been accomplished, from there on to avoid completely any mentioning of the regular P.B. with their confusing independent differentiations with respect to q's and p's which by (9.13) we want to consider not really independent.

Since the old modernized regular P.B. appeared not only in the definition (9.4) of the modified P.B., but also in the equations of motion of Hamilton (7.5), we want to express these equations of motion now in terms of modified P.B. With the covariant assumption (8.10a), the difference between $[F,\mathcal{H}_o]$ and $(F,\mathcal{H}_o)$ by (9.4) is seen to vanish strongly on account of (8.10b). Thus we find

$$F_{,0} \equiv [F,\mathcal{H}_o] + o^i(\partial^A F/\partial p^i) + o^i{}_{,r}(\partial^A F/\partial p^i{}_{,r}) \quad (9.15)$$

for any function F of the q, $q_{,s}$, p, and $p_{,r}$. Among the o^i appearing in (9.15), now, the o_n will strongly vanish on account of (7.3a) with (8.10a) and the time derivative of (9.13):

$$o_n \equiv \theta_{n,0} - \zeta_n \equiv 0; \quad (9.16a\text{–}b)$$

however, the other o^i in general will not vanish strongly. Since the $\eta^i{}_{,0}$ in general do not transform contragrediently to q_i, any imposed condition that the o^i given by (7.3a) would strongly vanish would not even seem to be a covariant condition. The reason why the conditions $o_n \equiv 0$ were covariant is because in (9.16) the $\theta_{n,0}$ and the ζ_n separately vanish strongly, and because, besides the $\theta_{n,0}$, also the $\theta_{n,s}$ with $s=1$, 2, 3 vanish strongly on account of (9.13), so that the $\theta_{n,0}$ remain zero after a coordinate transformation mixing time differentiations with spatial differentiations. If the same would have to be true for the other o^i, one would expect a strong vanishing of the η^i in analogy to (9.13); but the original reasons for which we introduced the η^i make this impossible. Therefore, $F_{,0}$ equals $[F,\mathcal{H}_o]$ only *weakly*, on account of (7.3b). This point will be important in Sec. 13 when we discuss the question of the covariance of our quantization.

Now using the expression (6.9) for $\mathcal{H}_o$, and using the fact that the Φ^k and any modified P.B. of O_2 with another quantity vanish at least weakly, and that the θ_n and its modified P.B. with other quantities vanish strongly, we obtain the weak equation of motion

$$F_{,0} = [F,\mathcal{H}_o] = \int d^3\mathbf{x}'\{[F,\mathfrak{H}(\mathbf{x}')] + [F,\Phi^k(\mathbf{x}')]B_k(\mathbf{x}')\}. \quad (9.17)$$

This looks like an equation of motion for F, but yet it is not always. For instance, if one chooses $F \equiv A_0$, then explicit calculation of the right-hand member of (9.17) yields

$$4\pi(g^{00}j)^{-1}\, P^\alpha g_{\alpha 0} + (g^{00})^{-1}\, g^{r0}(A_{r,0} - A_{0,r}) + A_{0,0}, \quad (9.18)$$

and the last term of this cancels against the $A_{0,0}$ in the left-hand member of (9.17), and one winds up not with an equation of motion for A_0, but with an equation for $g^{r0}A_{r,0}$.

We have been working here in a classical approximation only. In the quantum-mechanical generalization of (9.17), the modified P.B. are replaced by $(i\hbar c)^{-1}\times$ the corresponding commutator, hoping for the best with respect to the assumption (4.6) various times made tacitly in the course of the derivation. A check on this assumption is not possible as long as we do not specify our η^i in terms of the q and p. Moreover, in the "true theory" introduced in Sec. 24, a check of this sort becomes superfluous. Finally, for reasons discussed in Sec. 12, a change from modified P.B. to commutators in the third member of (9.17) with use of the expressions for $\mathfrak{H}$, etc. derived in the foregoing cannot be expected to lead to correct quantum-mechanical equations of motion anyhow.

10. COMMUTATION RELATIONS OF THE ORIGINAL VARIABLES

We shall now calculate the modified P.B. between the various q and p and such derived variables as $g_{\mu\nu}$ or $j\equiv\sqrt{-g}$. According to (9.3), the commutation (or anticommutation) relations between these variables are then obtained by multiplying the modified P.B. by $i\hbar c$. The (anti)commutators of more complicated expressions then should be obtained from the ones derived here for the simple variables, even if the result thus obtained might differ from what one would obtain by first calculating the modified P.B. between the complicated expressions and then multiplying by $(i\hbar c)$. The existence of such differences for complicated expressions A and B was the reason for writing "$\lim(\hbar\rightarrow 0)$" in front of (9.3).

Using (9.4) with the θ_n given by (6.4) with (5.1) and the C_{mn} given by (9.5) with (9.1), so that c^{lm} is given by

$$c_L{}^M \equiv c^M{}_L \equiv (\hbar c j g^{00})^{-1}(\gamma^0)_L{}^M\, \delta_3(\mathbf{x}''-\mathbf{x}'''), \qquad (10.1)$$

we find the following modernized modified Poisson brackets:

$$[h_{(\alpha)}{}^\mu, p_\nu{}^{(\beta)}] \equiv -[p_\nu{}^{(\beta)}, h_{(\alpha)}{}^\mu] \equiv \boldsymbol{\delta}_3\, \delta_\nu{}^\mu\, \delta_{(\alpha)}{}^{(\beta)}; \qquad (10.2a)$$

$$[p_\mu{}^{(\alpha)}, p_\nu{}^{(\beta)}] \equiv \tfrac{1}{4}\hbar c\, \boldsymbol{\delta}_3\, \delta_\mu{}^0\, \delta_\nu{}^0 (g^{00})^{-1}\, j \times\bar{\psi}[\gamma^{(\alpha)}\gamma^0\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^0\gamma^{(\alpha)}]\psi; \qquad (10.2b)$$

$$[p_\mu{}^{(\alpha)}, \psi_K] \equiv -[\psi_K, p_\mu{}^{(\alpha)}] \equiv \tfrac{1}{2}\boldsymbol{\delta}_3\{(g^{00})^{-1}\delta_\mu{}^0(\gamma^0\gamma^{(\alpha)}\psi)_K - h^{(\alpha)}{}_\mu\,\psi_K\}, \qquad (10.2c)$$

$$[p_\mu{}^{(\alpha)}, \bar{\psi}^L] \equiv -[\bar{\psi}^L, p_\mu{}^{(\alpha)}] \equiv \tfrac{1}{2}\boldsymbol{\delta}_3\{(g^{00})^{-1}\delta_\mu{}^0(\bar{\psi}\gamma^{(\alpha)}\gamma^0)^L - \bar{\psi}^L\, h^{(\alpha)}{}_\mu\}; \qquad (10.2d)$$

$$[\psi_K, \bar{\psi}^L] \equiv +[\bar{\psi}^L, \psi_K] \equiv -(\hbar c g^{00} j)^{-1}\, \boldsymbol{\delta}_3\, (\gamma^0)_K{}^L; \qquad (10.2e)$$

$$[A_\mu, P^\nu] \equiv -[P^\nu, A_\mu] \equiv \boldsymbol{\delta}_3\, \delta_\mu{}^\nu. \qquad (10.2f)$$

Here, $\boldsymbol{\delta}_3$ denotes the three-dimensional delta function between the two different points at which the two quantities are taken of which we are giving the modified P.B. Other pairs among the variables $h_{(\alpha)}{}^\mu$, $p_\nu{}^{(\beta)}$, ψ_K, $\bar{\psi}^L$, A_μ, and P^ν all have vanishing modified P.B. The modified P.B. for π^K and for π_L can be calculated by (5.1a) and (5.1b) with $\bar{\eta}\equiv 0$, $\eta\equiv 0$ from those for $\bar{\psi}^L$ and ψ_K:

$$[p_\mu{}^{(\alpha)}, \pi_K] \equiv -[\pi_K, p_\mu{}^{(\alpha)}] \equiv \tfrac{1}{4}\hbar c\, j\, \boldsymbol{\delta}_3\, \{[\delta_\mu{}^0\,\gamma^{(\alpha)} - h^{(\alpha)}{}_\mu\,\gamma^0]\psi\}_K, \qquad (10.3a)$$

$$[p_\mu{}^{(\alpha)}, \bar{\pi}^L] \equiv -[\bar{\pi}^L, p_\mu{}^{(\alpha)}] \equiv \tfrac{1}{4}\hbar c\, j\, \boldsymbol{\delta}_3\, \{\bar{\psi}[\delta_\mu{}^0\,\gamma^{(\alpha)} - h^{(\alpha)}{}_\mu\,\gamma^0]\}^L; \qquad (10.3b)$$

$$[\psi_K, \bar{\pi}^L] \equiv [\bar{\pi}^L, \psi_K] \equiv [\pi_K, \bar{\psi}^L] \equiv [\bar{\psi}^L, \pi_K] \equiv \tfrac{1}{2}\boldsymbol{\delta}_3\delta_K{}^L; \qquad (10.3c)$$

$$[\pi_K, \bar{\pi}^L] \equiv [\bar{\pi}^L, \pi_K] \equiv -\tfrac{1}{4}\hbar c(\gamma^0)_K{}^L\, \boldsymbol{\delta}_3. \qquad (10.3d)$$

We further give some useful modified P.B. of derived quantities:

$$[p_\mu{}^{(\alpha)}, g_{\rho\sigma}] \equiv -[g_{\rho\sigma}, p_\mu{}^{(\alpha)}] \equiv (g_{\mu\rho}h^{(\alpha)}{}_\sigma + g_{\mu\sigma}h^{(\alpha)}{}_\rho)\boldsymbol{\delta}_3; \qquad (10.4a)$$

$$[g^{\rho\sigma}, p_\mu{}^{(\alpha)}] \equiv -[p_\mu{}^{(\alpha)}, g^{\rho\sigma}] \equiv (\delta_\mu{}^\rho\, h^{(\alpha)\sigma} + \delta_\mu{}^\sigma\, h^{(\alpha)\rho})\boldsymbol{\delta}_3; \qquad (10.4b)$$

$$[p_\mu{}^{(\alpha)}, j] \equiv -[j, p_\mu{}^{(\alpha)}] \equiv j\, h^{(\alpha)}{}_\mu\, \boldsymbol{\delta}_3. \qquad (10.4c)$$

The "commutators" (2.5) for all these pairs of quantities are simply $i\hbar c\times$ the modified P.B. given above.

11. CONSISTENCY OF THE CONSTRAINTS

As the various constraint functions Φ^k are weakly to vanish simultaneously, also their products and therefore their commutators and anticommutators should vanish at least weakly. For compatibility with the quantization prescription (9.3) therefore it is a necessary condition that all the modified P.B. of the Φ^k among each other vanish at least weakly. By explicit calculation one finds

$$[\Phi^{(\alpha)(\beta)}, \Phi^{(\gamma)(\delta)\prime}] \equiv \{g^{(\alpha)(\delta)}\Phi^{(\beta)(\gamma)} + g^{(\beta)(\gamma)}\Phi^{(\alpha)(\delta)} + g^{(\gamma)(\alpha)}\Phi^{(\delta)(\beta)} + g^{(\delta)(\beta)}\Phi^{(\gamma)(\alpha)}\}\boldsymbol{\delta}_3, \qquad (11.1a)$$

$$[\Phi_\mu, \Phi_\nu{}'] \equiv \{\Phi_\mu\delta_\nu{}^0 - \Phi_\nu\delta_\mu{}^0\}\boldsymbol{\delta}_3, \qquad (11.1b)$$

$$[\Phi, \Phi'] \equiv [\Phi, \Phi_\mu{}'] \equiv [\Phi, \Phi^{(\alpha)(\beta)\prime}] \equiv [\Phi_\mu, \Phi^{(\alpha)(\beta)\prime}] \equiv 0, \qquad (11.1\text{c–f})$$

so that, indeed, the modified P.B. of these constraints will vanish weakly on account of these constraints.

As explained above (8.1a), weakly

$$\Phi^k{}_{,0} = 0. \qquad (11.2)$$

We evaluate $\Phi^k{}_{,0}$ weakly by (9.17) with F replaced by Φ^k. The last term in (9.17) will vanish weakly by (11.1) with the assumption (4.6). Therefore, the weak equations

$$\int d^3\mathbf{x}'\, [\Phi^k, \mathfrak{H}(\mathbf{x}')] = 0 \qquad (11.3)$$

must be valid. Explicit calculation of the left-hand members of the equations (11.3) yields[32]

$$\int d^3\mathbf{x}'\, [\Phi, \mathfrak{H}'] \equiv \Omega + \chi, \qquad (11.4a)$$

$$\int d^3\mathbf{x}'\, [\Phi_\mu, \mathfrak{H}'] \equiv \Omega_\mu + A_\mu\chi + \chi_\mu, \qquad (11.4b)$$

$$\int d^3\mathbf{x}'\, [\Phi^{(\alpha)(\beta)}, \mathfrak{H}'] \equiv 0, \qquad (11.4c)$$

where χ and χ_μ are the expressions (8.3), and where Ω and Ω_μ are the expressions

$$\Omega \equiv -[(g^{00})^{-1} g^{0s} \Phi]_{,s}, \tag{11.5a}$$

$$\begin{aligned}\Omega_\mu \equiv \tfrac{1}{2}\delta_\mu{}^s h_{(\beta)}{}^\lambda{}_{,s} h_{(\alpha)\lambda}\Phi^{(\alpha)(\beta)} - 4\pi(g^{00}j)^{-1} g_{\lambda\mu}P^\lambda\Phi \\ + [(g^{00})^{-1}\{h_{(\alpha)\mu}(\tfrac{1}{2}h_{(\beta)}{}^s g^{00} - h_{(\beta)}{}^0 g^{0s})\Phi^{(\alpha)(\beta)} \\ - 2g^{0s}\Phi_\mu\}]_{,s}.\end{aligned} \tag{11.5b}$$

Since the latter expressions vanish weakly on account of the primary constraints, we conclude from (11.4) that the expressions χ and χ_μ must vanish weakly. They are functions of the q, $q_{,r}$, $q_{,rs}$, p, and $p_{,s}$. Thus we have found the secondary constraints (8.2) to be additional necessary conditions for the consistency of our theory.

Consistency now requires that also the products and the "commutators" and therefore the modified P.B. of the functions χ^k among each other as well as with the functions Φ^k shall vanish at least weakly. Explicit calculation of these modified P.B. yields[32]

$$[\chi, \Phi'] \equiv [\chi, \Phi_\mu'] \equiv [\chi, \Phi^{(\alpha)(\beta)\prime}] \equiv 0; \tag{11.6a–c}$$

$$[\chi_\mu, \Phi'] \equiv -\delta_\mu{}^0\chi\dot{\boldsymbol{\delta}}_3 + (g^{00})^{-1}(g^{00}\delta_\mu{}^s - g^{0s}\delta_\mu{}^0)\Phi\boldsymbol{\delta}_{3,s}, \tag{11.6d}$$

$$\begin{aligned}[\chi_\mu, \Phi_\nu'] \equiv [\Phi_\mu\delta_\nu{}^s\boldsymbol{\delta}_3]_{,s} \\ + \Phi_\nu(g^{00})^{-1}(g^{00}\delta_\mu{}^s - 2\delta_\mu{}^0 g^{0s})\boldsymbol{\delta}_{3,s} \\ + \delta_\mu{}^0\Phi^{(\alpha)(\beta)}(g^{00})^{-1}(\tfrac{1}{2}g^{00}h_{(\beta)}{}^s - g^{0s}h_{(\beta)}{}^0)h_{(\alpha)\nu}\boldsymbol{\delta}_{3,s} \\ + \delta_\mu{}^0[4\pi(g^{00}j)^{-1}g_{\mu\nu}P^\nu\Phi \\ + \tfrac{1}{2}\delta_\nu{}^s h_{(\alpha)}{}^\lambda{}_{,s} h_{(\beta)\lambda}\Phi^{(\alpha)(\beta)} - \chi_\nu - A_\nu\chi]\boldsymbol{\delta}_3,\end{aligned} \tag{11.6e}$$

$$[\chi_\mu, \Phi^{(\alpha)(\beta)\prime}] \equiv \delta_\mu{}^s \Phi^{(\alpha)(\beta)} \boldsymbol{\delta}_{3,s}; \tag{11.6f}$$

$$[\chi, \chi'] \equiv 0, \quad [\chi_\mu, \chi'] \equiv \delta_\mu{}^s\chi\boldsymbol{\delta}_{3,s}, \tag{11.7a–b}$$

$$\begin{aligned}[\chi_\mu, \chi_\nu'] \equiv (\chi_\nu\delta_\mu{}^r + \Phi Z^r\delta_\mu{}^0\delta_\nu{}^0)\boldsymbol{\delta}_{3,r} - [\Phi_\mu{}^{rs}\delta_\nu{}^0\boldsymbol{\delta}_3]_{,rs} \\ + [(\chi_\mu\delta_\nu{}^r + \Phi Z^r\delta_\mu{}^0\delta_\nu{}^0)\boldsymbol{\delta}_3]_{,r} + \Phi_\nu{}^{rs}\delta_\mu{}^0\boldsymbol{\delta}_{3,rs}.\end{aligned} \tag{11.7c}$$

Here we put

$$Z^r \equiv (g^{00})^{-2}[A_{\lambda,s}(g^{0\lambda}g^{rs} - g^{0s}g^{r\lambda}) - 4\pi j^{-1}P^r], \tag{11.8a}$$

$$\Phi_\mu{}^{rs} \equiv (g^{00})^{-1}g^{rs}[\Phi_\mu + \tfrac{1}{2}h_{(\alpha)\mu}h_{(\beta)}{}^0\Phi^{(\alpha)(\beta)}]. \tag{11.8b}$$

Obviously the expressions (11.6)–(11.7) all vanish weakly on account of the primary and secondary constraints and the assumption (4.6), without imposing further independent constraints.

Finally, since the χ constraints can be derived for any value of x^0, we conclude that necessarily

$$\chi^k{}_{,0} = 0 \tag{11.9}$$

and, therefore, by (9.17) and (11.6),

$$\int d^3x' \, [\chi^k(\mathbf{x}), \mathfrak{H}(\mathbf{x}')] = 0. \tag{11.10}$$

Explicit calculation of the left-hand members of (11.10) yields[32]

$$[\chi, \int \mathfrak{H}' d^3x'] = 0, \tag{11.11a}$$

$$[\chi_\mu, \int \mathfrak{H}' d^3x'] = \delta_\mu{}^0[\Phi Z^r]_{,r} - \Phi_\mu{}^{rs}{}_{,rs}. \tag{11.11b}$$

This shows that the necessary conditions (11.10) are satisfied automatically and do not lead to further constraints.

Because of the relations (11.1), (11.4), (11.6), (11.7), and (11.11), our Φ constraints and χ constraints form a system of *first-class* constraints in Dirac's sense.[5,31] This completes the proof of the consistency of the canonical formulation of our theory[22 35]

12. CANONIZATION

From (9.3) with (10.1)–(10.3) we see that, as a consequence of the "modification" procedure (9.4), the pairs of variables $h_{(\alpha)}{}^\mu$ with $p_\mu{}^{(\alpha)}$ and ψ_K with $\bar{\pi}^K$ or $\bar{\psi}^L$ with π_L do not satisfy the conventional canonical commutation relations (2.1). We shall now introduce new variables which will satisfy the conventional commutation relations (2.1) on account of our modified quantization (9.3). That is, the modified P.B. of our new variables $\tilde{q}$ and $\tilde{p}$ will be

$$[\tilde{q}_i(\mathbf{x},x^0), \tilde{p}^j(\mathbf{x}',x^0)] = \delta_i{}^j\boldsymbol{\delta}_3(\mathbf{x} - \mathbf{x}'), \tag{12.1a}$$

$$[\tilde{q}_i(\mathbf{x},x^0), \tilde{q}_j(\mathbf{x}',x^0)] = [\tilde{p}^i(\mathbf{x},x^0), \tilde{p}^j(\mathbf{x}',x^0)] = 0. \tag{12.1b–c}$$

Now, let (A,B') denote the "canonized" P.B.,[35a] that is, modernized regular P.B. obtained from (2.10) by replacing in it the sums over derivatives with respect to the original q and p, by sums over derivatives with respect to the new "canonical" variables $\tilde{q}$ and $\tilde{p}$. Then, from (12.1) with (9.3) one finds for the modified P.B. of such functions A and B' of the canonical variables and their spatial derivatives in $\mathbf{x}$ and in $\mathbf{x}'$, in the classical limit (9.10) in which relations like (2.8) hold for modified P.B.,

$$\lim_{\hbar\to 0}[A,B'] = \lim_{\hbar\to 0}(A,B') = \lim_{\hbar\to 0}\frac{[A;B']}{i\hbar c}. \tag{12.2a–b}$$

From the new variables $\tilde{q}$ and $\tilde{p}$ we also require that the new variables $\tilde{p}_\mu{}^{(\alpha)}$ shall be Hermitian, and, for ensuring the validity of the conclusions conventionally drawn from the theory of second quantization,[36] that the canonical conjugates to the q describing the electron field shall differ from their Hermitian conjugates by a factor $i\hbar c$ only.[9]

This problem of "canonizing" the quantum-mechanical theory by introduction of appropriate new variables was first solved by the DeWitts.[6] Independently, the same method in a more explicit form was later used by Belinfante and Swihart in their linear

[31] We say "in Dirac's sense" and not "according to Dirac's formalism", because Dirac[5] used regular P.B. in the discussion of whether a set of constraints is first class or not, while we use (modernized) modified P.B., identifying the latter with commutators by (9.3). Therefore, we are not interested in the regular P.B. of the constraints, unless the latter are understood in the way discussed in Sec. 12.

[35] J. L. Anderson, Phys. Rev. 99, 1009 (1955).

[35a] Notice that we use here boldface parentheses, for distinction from the lightface parentheses used in (2.9)–(2.11).

[36] V. Fock, Z. Physik 75, 622 (1932); 76, 852 (1932).

theory of gravitation.[7,8] We use here a notation mostly resembling the latter work.[37]

We start from the expression for the charge density in units e:

$$\mathfrak{j}^0/e \equiv i\, j\, \bar\psi\gamma^0\psi \equiv \psi^\dagger\, j\, h_{(\alpha)}{}^0\, \alpha^{(\alpha)}\psi, \quad (12.3\text{a–b})$$

[compare (3.4) and (2.19)–(2.22)]. We now determine a Hermitian operator Θ which is one of the square roots of the operator $j\, h_{(\alpha)}{}^0 \alpha^{(\alpha)}$ appearing in (12.3):

$$\Theta^2 \equiv j\, h_{(\alpha)}{}^0 \alpha^{(\alpha)}, \quad \Theta \equiv \Theta^\dagger. \quad (12.4\text{a–b})$$

For this purpose and for future use, we introduce the notation[37]

$$\zeta \equiv F_{(0)} \equiv -F^{(0)} \equiv h_{(0)}{}^0\, j \equiv -h^{(0)0}\sqrt{-g}, \quad (12.5\text{a})$$

$$F^{(k)} \equiv F_{(k)} \equiv h_{(k)}{}^0 j \equiv h^{(k)0}\sqrt{-g}, \quad (12.5\text{b})$$

$$F \equiv |\mathbf{F}| \equiv (F_{(1)}{}^2 + F_{(2)}{}^2 + F_{(3)}{}^2)^{\frac12}, \quad (12.5\text{c})$$

$$v \equiv +(\zeta - F)^{\frac12}, \quad w \equiv +(\zeta + F)^{\frac12}, \quad (12.5\text{d})$$

$$\vartheta \equiv vw \equiv +(\zeta^2 - F^2)^{\frac12}, \quad \boldsymbol{\lambda} \equiv \mathbf{F}/F, \quad (12.5\text{e–f})$$

so that

$$F \equiv +(\zeta^2 - \vartheta^2)^{\frac12}, \quad (12.6\text{a})$$

$$\lambda_{(k)} \equiv h_{(k)}{}^0\, j\, (\zeta^2 - \vartheta^2)^{-\frac12}, \quad (12.6\text{b})$$

$$\vartheta^2 \equiv -F_{(\alpha)}F^{(\alpha)} \equiv (-g)(-g^{00}) \equiv -j\mathfrak{g}^{00}, \quad (12.6\text{c})$$

$$w \pm v \equiv +[2(\zeta \pm \vartheta)]^{\frac12}. \quad (12.6\text{d})$$

Using this notation, we find[7,8,37]

$$\Theta^2 \equiv \zeta + \boldsymbol{\alpha}\cdot\mathbf{F}, \quad (12.4\text{c})$$

$$\Theta \equiv \tfrac12(w+v) + \tfrac12(w-v)\boldsymbol{\lambda}\cdot\boldsymbol{\alpha} \equiv \Theta^\dagger$$
$$\equiv 2^{-\frac12}[(\zeta+\vartheta)^{\frac12} + (\zeta-\vartheta)^{\frac12}\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}], \quad (12.7)$$

$$\Theta^n \equiv \tfrac12 w^n[1+\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}] + \tfrac12 v^n[1-\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}], \quad (12.8\text{a})$$

$$f(\Theta) \equiv \tfrac12 f(w)[1+\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}] + \tfrac12 f(v)[1-\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}], \quad (12.8\text{b})$$

$$\Theta^{-1} \equiv (2\vartheta)^{-1}[(v+w) + (v-w)\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}]$$
$$\equiv 2^{-\frac12}\vartheta^{-1}[(\zeta+\vartheta)^{\frac12} - (\zeta-\vartheta)^{\frac12}\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}]. \quad (12.8\text{c})$$

We may also write

$$\Theta \equiv \frac{(-g)^{\frac14}\{(-g^{00})^{\frac12} - \gamma^{(0)}\gamma^0\}}{2^{\frac12}[(-g^{00})^{\frac12} + h_{(0)}{}^0]^{\frac12}}, \quad (12.9)$$

$$\Theta^{-1} \equiv \frac{1-(-g^{00})^{-\frac12}\gamma^0\gamma^{(0)}}{2^{\frac12}(-g)^{\frac14}[(-g^{00})^{\frac12} + h_{(0)}{}^0]^{\frac12}}. \quad (12.10)$$

With the help of the Hermitian Dirac matrix Θ given by (12.7) or (12.9) we now define new fields $\tilde\psi$ and $\tilde\psi^\dagger$ by

$$\tilde\psi \equiv \Theta\psi, \quad \tilde\psi^\dagger \equiv \psi^\dagger\Theta \equiv \bar\psi\beta\Theta, \quad (12.11\text{a–b})$$

so that

$$\psi \equiv \Theta^{-1}\tilde\psi, \quad \bar\psi \equiv \tilde\psi^\dagger\Theta^{-1}\beta \equiv (g^{00}g)^{-\frac12}\tilde\psi^\dagger\beta\Theta, \quad (12.12\text{a–c})$$

where we used $\beta \equiv i\gamma^{(0)}$.

We now can eliminate ψ and $\bar\psi$ and their derivatives completely from the Lagrangian by substitution of (12.12), in which Θ by (12.9) is a given function of the fourleg components and the constant Dirac matrices $\gamma^{(\alpha)}$. Thus, $L_\circ$ becomes a function of $\tilde\psi$, $\tilde\psi^\dagger$, $h_{(\alpha)}{}^\mu$, A_μ, and their derivatives. As (12.4a) yields

$$\Theta^{-1}\, j\, \alpha^0\, \Theta^{-1} \equiv \Theta^{-1}\Theta^2\Theta^{-1} \equiv 1, \quad (12.13)$$

the terms with $\tilde\psi_{,0}$ and with $\tilde\psi^\dagger{}_{,0}$ in the Lagrangian density are then easily found to be

$$\tfrac12 i\hbar c[\tilde\psi^\dagger\tilde\psi_{,0} - \tilde\psi^\dagger{}_{,0}\tilde\psi]. \quad (12.14)$$

Adding $\frac12 i\hbar c(\tilde\psi^\dagger\tilde\psi)_{,0}$ to the Lagrangian density, we get a new but equivalent Lagrangian with density

$$\tilde L_\circ(\tilde q, \tilde q_{,\mu}) = L_\circ(q, q_{,\mu}) + \tfrac12 i\hbar c(\tilde\psi^\dagger\tilde\psi)_{,0} \quad (12.15)$$

which is independent of $\tilde\psi^\dagger{}_{,0}$ and which contains $\tilde\psi_{,0}$ only in the form

$$i\hbar c\tilde\psi^\dagger\tilde\psi_{,0}. \quad (12.16)$$

We now consider $\tilde\psi$, $h_{(\alpha)}{}^\mu$, and A_μ as the variables $\tilde q_{\mathrm i}$, and we define the variables $\tilde p^{\mathrm i}$ by

$$\tilde p^{\mathrm i} \equiv \eta^{\mathrm i} + \partial^P\tilde L_\circ/\partial\tilde q_{\mathrm i,0}. \quad (12.17)$$

Then, the $\tilde p^{\mathrm i}$ canonically conjugate to

$$\tilde q_{\mathrm i} \equiv A_\mu, \quad \tilde\psi, \quad \text{and} \quad h_{(\alpha)}{}^\mu \quad (12.18\text{a})$$

are found to be

$$\tilde p^{\mathrm i} \equiv P^\mu, \; i\hbar c\tilde\psi^\dagger, \; \text{and} \; \tilde p_\mu{}^{(\alpha)}, \quad (12.18\text{b})$$

where $\tilde p_\mu{}^{(\alpha)}$ differs from $p_\mu{}^{(\alpha)}$ as we keep $\tilde\psi$ instead of ψ constant in the differentiation with respect to $h_{(\alpha)}{}^\mu{}_{,0}$. By explicit calculation we find

$$\tilde p_\mu{}^{(\alpha)} \equiv p_\mu{}^{(\alpha)} + \tfrac12 i\hbar c\tilde\psi^\dagger\left[\Theta^{-1}\frac{\partial\Theta_{,0}}{\partial h_{(\alpha)}{}^\mu{}_{,0}} - \frac{\partial\Theta_{,0}}{\partial h_{(\alpha)}{}^\mu{}_{,0}}\Theta^{-1}\right]\tilde\psi$$

$$\equiv p_\mu{}^{(\alpha)} + \tfrac12 i\hbar c\psi^\dagger\left[\frac{\partial\Theta}{\partial h_{(\alpha)}{}^\mu}\Theta - \Theta\frac{\partial\Theta}{\partial h_{(\alpha)}{}^\mu}\right]\psi$$

$$\equiv p_\mu{}^{(\alpha)} + \frac{\hbar c j\delta_\mu{}^0\, h_{(\beta)}{}^0\, \bar\psi[\gamma^{(\alpha)}\gamma^{(0)}\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^{(0)}\gamma^{(\alpha)}]\psi}{4[h_{(0)}{}^0 + (-g^{00})^{\frac12}]}, \quad (12.19)$$

or, conversely,

$$p_\mu{}^{(\alpha)} \equiv \tilde p_\mu{}^{(\alpha)} + \frac{\hbar c\delta_\mu{}^0\tilde\psi^\dagger\beta[\gamma^{(\alpha)}\gamma^0\gamma^{(0)} - \gamma^{(0)}\gamma^0\gamma^{(\alpha)}]\tilde\psi}{4[h_{(0)}{}^0(-g^{00})^{\frac12} - g^{00}]}. \quad (12.20)$$

In the notation of (12.5) and (2.20), (12.19) takes the form

$$\tilde p_\mu{}^{(0)} \equiv p_\mu{}^{(0)}, \quad \tilde p_s{}^{(\alpha)} \equiv p_s{}^{(\alpha)}, \quad (12.21\text{a–b})$$

[37] Our present $\mathbf{F}$, F, $\boldsymbol{\lambda}$, and $\tilde\psi$ correspond to $-\frac12\boldsymbol{\Gamma}$, $+\frac12\Gamma$, $-\mathbf{t}$, and φ of references 7 and 8. In reference 7, D was written for our present ϑ. The relation between the quantities ζ and $\mathbf{F}$ on the one hand, and the gravitational field on the other hand, in our present work differs, of course, from the corresponding relation in the linear theory of gravitation.

$$\tilde{p}_0^{(l)} \equiv p_0^{(l)} + \tfrac{1}{2}\hbar c(\zeta + \vartheta)^{-1} j\, \epsilon^{(l)(m)(n)} (\psi^\dagger \sigma_{(m)}\psi) F_{(n)}, \quad (12.21c)$$

where $\epsilon^{(l)(m)(n)} = 1, -1$, or 0, as $(l)(m)(n)$ is an even, an odd, or no permutation at all of (1)(2)(3).

Using the modified P.B. (10.2), it can then be shown by explicit calculation that the $\tilde{q}$ and $\tilde{p}$ thus introduced as functions of the old q and p automatically satisfy the relations (12.1) and therefore are the desired canonical variables, quantized according to (2.1) with symbols $\sim$ placed over the q and p. In the verification that $[\tilde{p}_\mu^{(\alpha)}(\mathbf{x}), \tilde{p}_\nu^{(\beta)}(\mathbf{x}')] \equiv 0$, use is made of the identity[7]

$$X_{(k)}\epsilon_{(l)(m)(n)} + X_{(l)}\epsilon_{(m)(k)(n)} + X_{(m)}\epsilon_{(k)(l)(n)} + X_{(n)}\epsilon_{(m)(l)(k)} \equiv 0, \quad (12.22)$$

which is valid for *any* quantity $X_{(k)}$, as long as the indices take the values (1), (2), (3) only. In particular, one applies (12.22) with $X_{(k)} \equiv F_{(k)}$ [see (12.5b)].

The charge density, in terms of the new variables, is given by

$$\rho \equiv \mathrm{j}^0 \equiv e\psi^\dagger \Theta^2 \psi \equiv e\tilde{\psi}^\dagger \tilde{\psi}. \quad (12.23)$$

Since $i\hbar c\tilde{\psi}^\dagger$ is now the canonical conjugate to $\tilde{\psi}$, the $\tilde{\psi}$-field is quantized according to[38]

$$\tilde{\psi}_K(\mathbf{x})\tilde{\psi}^{*L}(\mathbf{x}') + \tilde{\psi}^{*L}(\mathbf{x}')\tilde{\psi}_K(\mathbf{x}) \equiv \delta_K{}^L\, \delta_3(\mathbf{x}-\mathbf{x}'), \quad (12.24a)$$

$$\tilde{\psi}_K\tilde{\psi}_L' + \tilde{\psi}_L'\tilde{\psi}_K \equiv 0, \quad (12.24b)$$

$$\tilde{\psi}^{*K}\tilde{\psi}^{*L'} + \tilde{\psi}^{*L'}\tilde{\psi}^{*K} \equiv 0, \quad (12.24c)$$

so that the formulas of the conventional theory of second quantization[36] are valid without change. In particular, if Q_V is the charge at time x^0 in a volume V,

$$Q_V(x^0) \equiv \int_V \rho(\mathbf{x},x^0)\, d^3\mathbf{x} \equiv e\int_V d^3\mathbf{x}\, \tilde{\psi}^{*L}\tilde{\psi}_L, \quad (12.25)$$

and if $\delta(\mathbf{x},V) = 1$ for $\mathbf{x}$ inside V, and $\delta(\mathbf{x},V) = 0$ for $\mathbf{x}$ outside V, then (12.24)–(12.25) yield

$$Q_V\, \tilde{\psi}_K(\mathbf{x}) \equiv \tilde{\psi}_K(\mathbf{x})[Q_V - e\delta(\mathbf{x},V)], \quad (12.26a)$$

$$Q_V\, \tilde{\psi}^{*K}(\mathbf{x}) \equiv \tilde{\psi}^{*K}(\mathbf{x})[Q_V + e\delta(\mathbf{x},V)]. \quad (12.26b)$$

From this it follows that $\tilde{\psi}_K(\mathbf{x})$ and $\tilde{\psi}^{*K}(\mathbf{x})$ are operators which decrease and increase the charge Q_V in any volume V containing $\mathbf{x}$ by e, while they do not change the charge in volumes not containing $\mathbf{x}$. Therefore, $\tilde{\psi}$ and its conjugate $\tilde{\psi}^\dagger$, and not the old undors ψ and $\bar{\psi}$, should be considered the "electron wave functions."

We see from (12.2b) that, if we had expressed the Lagrangian from the beginning in the form $\tilde{L}_\circ$, we would have had no need for introducing modified P.B. We can now derive a Hamiltonian $\tilde{\mathcal{H}}_\circ$ directly from $\tilde{L}_\circ$. In the quantum-mechanical theory, we should then "symmetrize" in $\tilde{\mathcal{H}}_\circ$ all terms according to the traditional prescription (6.13). This is preferable to first symmetrizing $\tilde{\mathcal{H}}_\circ$ by (6.13) and then eliminating the old variables by (12.12) and (12.20), because such a procedure could be expected to lead to Hamiltonian equations which are not equivalent to the Lagrangian equations because of noncommutativity of certain factors.[7–8]

[38] As by (2.20) we chose $\beta^2 = 1$, we may identify dotted undor subscripts with undotted undor superscripts, so that ψ^{*K} may be written instead of $\psi^*{}_{\dot{K}}$ for the Hermitian conjugate of the component ψ_K of ψ. For the $\tilde{\psi}$ and $\tilde{\psi}^\dagger$ we use a similar index notation, although these quantities are not undors at all under Lorentz transformations of the fourlegs.

13. DIFFICULTIES WITH RESPECT TO THE COVARIANCE

We now try to prove the covariance of our commutation relations for the canonical variables $\tilde{q}$ and $\tilde{p}$. It would be sufficient to prove that, whenever they are valid on one surface $x^0 =$ constant, then on this surface all their time derivatives would be valid. By Taylor expansion in x^0, their validity on *any* surface $x^0 =$ constant would then follow. This would then complete the proof of covariance of the commutation relations, on account of the arbitrariness of the coordinate system used.[39,40]

For the time derivative of the "commutator" $[\tilde{q}_{\mathrm{i}}\, ; \tilde{p}^{\mathrm{j}}]$ one easily finds, by a generalization of (9.15),

$$[\tilde{q}_{\mathrm{i}}\, ; \tilde{p}^{\mathrm{j}\prime}]_{,0} \equiv [[\tilde{q}_{\mathrm{i}}\, ; \tilde{p}^{\mathrm{j}\prime}], \tilde{\mathcal{H}}_\circ] + o^{\mathrm{k}}(\partial^A[\tilde{q}_{\mathrm{i}}\, ; \tilde{p}^{\mathrm{j}\prime}]/\partial p^{\mathrm{k}}) + o^{\mathrm{k}\prime}(\partial^A[\tilde{q}_{\mathrm{i}}\, ; \tilde{p}^{\mathrm{j}\prime}]/\partial p^{\mathrm{k}\prime}). \quad (13.1)$$

Remember the factor ordering discussed below (7.5). On the surface $x^0 =$ constant where we have already assumed the validity of

$$[\tilde{q}_{\mathrm{i}}\, ; \tilde{p}^{\mathrm{j}\prime}]_{(\mathrm{at}\ x^0)} \equiv i\hbar c[\tilde{q}_{\mathrm{i}}, \tilde{p}^{\mathrm{j}\prime}] \equiv i\hbar c\delta_{\mathrm{i}}{}^{\mathrm{j}}\, \delta_3(\mathbf{x}-\mathbf{x}'), \quad (13.2)$$

(13.1), by $[i\hbar c\delta_{\mathrm{i}}{}^{\mathrm{j}}\, \delta_3(\mathbf{x}-\mathbf{x}'), \mathcal{H}_\circ] \equiv 0$ and by (2.15a), simplifies to

$$\{[\tilde{q}_{\mathrm{i}}(\mathbf{x})\, ; \tilde{p}^{\mathrm{j}}(\mathbf{x}')]_{,0}\}_{(\mathrm{at}\ x^0)} \equiv \int d^3\mathbf{x}''\, o^{\mathrm{k}}(\mathbf{x}'')\{\eth^A[\tilde{q}_{\mathrm{i}}(\mathbf{x})\, ; \tilde{p}^{\mathrm{j}}(\mathbf{x}')]/\eth p^{\mathrm{k}}(\mathbf{x}'')\}. \quad (13.3)$$

The difficulty now is that, even though the commutator in the right-hand member of (13.3) is given by (13.2), yet the integrand in this member need not vanish. In this integrand, the order of factors is such as to substitute factors o^{k} for factors p^{k}, and such a substitution even on a vanishing commutator need not give a vanishing result, as is easily seen from the example

$$q_{\mathrm{i}}(\partial^A[p^{\mathrm{i}}\, ; p^{\mathrm{k}}]/\partial p^{\mathrm{k}}) \equiv [p^{\mathrm{i}}\, ; q_{\mathrm{i}}] \not\equiv 0, \quad (13.4)$$

with the same interpretation of the order of factors.

If all o^{k} were strongly zero, from the strong vanishing of (13.3) at x^0 would then follow the strong vanishing of its time derivative $[\tilde{q}_{\mathrm{i}}; \tilde{p}^{\mathrm{j}\prime}]_{,00}$ at x^0 by an argument similar to (13.1)–(13.3). By repeating the argument over and over again, we would then by Taylor expansion find (13.2) to be valid at any time. One could deal similarly with $[\tilde{q}\, ; \tilde{q}']$ and with $[\tilde{p}\, ; \tilde{p}']$.

[39] P. Weiss, Proc. Roy. Soc. (London) **A169**, 102 and 119 (1938).
[40] P. G. Bergmann and J. H. M. Brunings, Revs. Modern Phys. **21**, 480 (1949).

As explained in the text beneath (9.16), however, we cannot assume the o^k to vanish strongly. If, therefore, we use merely the *weak* vanishing of o^k, then, from the strong validity of (13.2) at the time x^0 only the weak vanishing of (13.3) would follow, so we cannot predict the value of $[\tilde{q};\tilde{p}']_{,00}$ at x^0 not knowing $[[\tilde{q}_i;\tilde{p}^{j\prime}]_{,0},\tilde{\mathcal{H}}_\circ]$. Thus, our proof of covariance completely breaks down.

These troubles arise from the fact that commutators of o^k with other quantities need not vanish strongly, and that commutators of weakly vanishing quantities with $\tilde{\mathcal{H}}_\circ$ need not vanish even weakly. The latter difficulty would be avoided if we could believe enough in the assumption (4.6). However, the use of (4.6) prior to (or for the purpose of) a calculation of a commutator would undo the entire introduction of weakly vanishing quantities η^i and ζ^i, of which the purpose was providing different commutation relations, even weakly, for instance for p^i and for $\partial^P L_\circ/\partial q_{i,0}$.

14. WHAT NOW?

The difficulties of the preceding section could perhaps be solved by a complete change of the quantization procedure, replacing (12.1) by commutation relations which would be compatible with strong vanishing of all η^i, ζ^i, and o^i. As attempts at the realization of such a program meet the unsolved problem of finding the so-called "true variables" of the theory,[41,42] we shall postpone its discussion to Sec. 24. We first consider a different attempt at escaping the difficulties.

The cause of the trouble is the unknown behavior of the o^k. The need for the o^k arose from the existence of constraints. We could avoid the latter if we would destroy the various invariance properties of the Lagrangian by adding noninvariant terms to $\tilde{L}_\circ$. Such noninvariant terms which were no part of the basic Lagrangian $\tilde{L}_\circ$ would not describe any physical reality. The theory using the Lagrangian including these noninvariant terms we shall call the altered theory, and we denote the added terms themselves by L_{add}. These terms really ought to vanish; we shall assume that they vanish "weakly." As these terms themselves are functions of the field variables and their derivatives, the condition of their weak vanishing imposes differential equations as restrictions on the field variables. These "auxiliary conditions" are not invariant under all transformations which left the basic Lagrangian invariant; the total altered Lagrangian will show invariance only under the more restricted group of transformations which leave the auxiliary conditions invariant. Because of the restrictions thus imposed upon the invariance group, the momenta defined by (3.1) in terms of the total altered Lagrangian will not be subject to constraints such as (3.8), (3.10), or (3.13).

The condition restricting the group of gauge transformations will essentially be the Lorentz condition; as a condition restricting the coordinate transformations we use the De Donder coordinate condition.[43] Finally, we impose a convenient fourleg condition. Let us represent all those auxiliary conditions by the weak equations

$$S^n(q,q_{,\mu}) = 0. \tag{14.1}$$

Then we may choose for our altered Lagrangian density

$$\tilde{L} \equiv \tilde{L}_\circ + L_{\text{add}} \tag{14.2a}$$

with

$$L_{\text{add}} \equiv K_{mn} S^m S^n, \tag{14.2b}$$

where for K_{mn} we choose a convenient symmetric set of coefficients

$$K_{mn} \equiv (-)^{u_m} K_{nm}, \tag{14.3}$$

with $u_m = u_n$. (In practice, all $u_m = 0$.)

We now define p^i by (3.1) with

$$L \equiv L_\circ + L_{\text{add}} \equiv \tilde{L} - \tfrac{1}{2} i\hbar c(\tilde{\psi}^\dagger\tilde{\psi})_{,0}. \tag{14.4}$$

Comparison with (4.3) then yields

$$\eta^i \equiv \partial^P L_{\text{add}}/\partial q_{i,0} \equiv K_{mn}\{S^m(\partial^P S^n/\partial q_{i,0}) + (-)^{u_n u_i}(\partial^P S^m/\partial q_{i,0})S^n\}. \tag{14.5a–b}$$

We can also write

$$\eta^i \equiv \partial^P L_{\text{add}}/\partial \tilde{q}_{i,0}. \tag{14.5c}$$

This is so because L_{add} will not depend on either $\tilde{\psi}$ or $\tilde{\psi}^\dagger$ or their derivatives, and those are the only $\tilde{q}$ which are not equal to the corresponding q. For these q^n, both (14.5a) and (14.5c) express the strong second-class constraints (9.13). [Compare (5.1).] From (14.5b) we also see that the weak vanishing of the other η^i, which was postulated in (4.3b), may now be regarded as a consequence of the weak conditions (14.1), provided that the assumption (4.6), that a weakly vanishing factor irrespective of its position makes a product weakly zero, is applicable to the last member of Eq. (14.5b).

By (14.5c) and (14.2a) we now find from (12.17)

$$\tilde{p}^i \equiv \partial^P \tilde{L}/\partial \tilde{q}_{i,0}, \tag{14.6}$$

so that our old $\tilde{q}$ and $\tilde{p}$ now are the canonical variables in our altered theory with Lagrangian density $\tilde{L}$.

We assume the strong validity of the altered Lagrangian field equations obtained from the variational principle

$$(\text{Boundary integral}) \equiv \delta\int \tilde{L}d^4x \equiv \delta\int L d^4x; \tag{14.7a–b}$$

thence,

$$\partial^P\tilde{L}/\partial\tilde{q}_i - (\partial^P\tilde{L}/\partial\tilde{q}_{i,\mu})_{,\mu} \equiv 0, \tag{14.8a}$$

$$\partial^P L/\partial q_i - (\partial^P L/\partial q_{i,\mu})_{,\mu} \equiv 0. \tag{14.8b}$$

Comparison with (4.4a) yields

$$\zeta^i \equiv (\partial^P L_{\text{add}}/\partial q_{i,\mu})_{,\mu} - \partial^P L_{\text{add}}/\partial q_i. \tag{14.9}$$

---

[41] P. G. Bergmann and I. Goldberg, Phys. Rev. 98, 2 and 544 (1955).

[42] P. G. Bergmann, Nuovo cimento 3, Ser. 10, 1177 (1956).

[43] T. De Donder, *La Gravifique Einsteinienne* (Gauthiers-Villars, Paris, 1921).

By (14.2b), this is a polynomial containing factors S^n or $S^n{}_{,\mu}$ in all of its terms, so that the weak vanishing (4.4b) of ζ^i is a consequence of (14.1) and its derivatives $S^n{}_{,\mu} = 0$, if again the assumption (4.6) is applicable.

Since by (4.7) we could express the original constraints in terms of the η, we see from (14.5b) that the constraints follow from the auxiliary conditions (14.1). Later, we complete the analogy between the constraints and the auxiliary conditions by showing that, conversely, the auxiliary conditions follow from the constraints. We shall also provide a simple proof, requiring much less explicit calculation than the derivation of (11.4c), (11.11a) and (11.11b) that, whenever (14.1) and some of their first time derivatives are valid at a given time x^0, automatically their higher time derivatives will be valid, and therefore (14.1) will hold for all times.[44] Repeatedly we use the assumption (4.6) that all expressions which we meet and which contain a factor S^n placed anywhere will vanish weakly on account of (14.1).

Equations (3.1) or (14.6), together with (14.8b) or (14.8a), enable us to express all the $q_{,0}$ and $\tilde{q}_{,0}$ in terms of the q and p or the $\tilde{q}$ and $\tilde{p}$ and their spatial derivatives. For instance, we find $\tilde{\psi}_{,0}$ from the field equation (14.8a) with $\tilde{q}_i \equiv \tilde{\psi}^\dagger$. The other $\tilde{q}_{,0}$ are solved for from (14.6), without any constraints interfering. Therefore, the canonical formalism in the altered theory takes the conventional form, using as the Hamiltonian the expression

$$\tilde{\mathcal{H}} \equiv \int \tilde{H} d^3\mathbf{x} = \tilde{\mathcal{H}}_\circ + \mathcal{H}_{\text{add}}, \tag{14.10a}$$

$$\tilde{H} = \tilde{p}^i \tilde{q}_{i,0} - \tilde{L} = \tilde{H}_\circ - L_{\text{add}} = \tilde{H}_\circ + H_{\text{add}}. \tag{14.10b}$$

Quantization then follows by Eqs. (12.1), while such expressions as (14.10) are subjected to the symmetrization prescription (6.13).

The o^i, which bothered us in the past, are now from (7.3), (14.5a), and (14.9) easily found to be equal to

$$o^i \equiv \partial^P L_{\text{add}}/\partial q_i - (\partial^P L_{\text{add}}/\partial q_{i,s})_{,s} \equiv -\mathfrak{d}^P \mathcal{H}_{\text{add}}/\mathfrak{d} q_i. \tag{14.11}$$

The main trouble with the altered theory is that its field equations are not strongly covariant, due to the terms (14.9) in (14.8b) or (4.4a) weakly vanishing on account of (4.4b) or (14.1). They could, however, even "weakly" cause troubles, if there would be reasons to disbelieve the assumption (4.6). Failure of (4.6) to be valid at various other places may cause further difficulties. It is hoped that such difficulties can be avoided by proceeding from the altered theory to the "true" theory introduced in Sec. 24.

In Lorentz-covariant quantum electrodynamics, both the altered theory and the true theory are well known. The altered theory is here Fermi's quantum electrodynamics,[45] which has formed the basis for most of the modern applications of quantum electrodynamics verified by experiments. Similarly, one might hope to find applications of the altered quantum theory of gravitation before the "true" theory is completely developed. However, in similarly altered quantum electrodynamics we had the advantage of knowing at least the "true variables," (there, the transverse fields), which for our present theory are not yet known explicitly. This may make it hard to formulate physically meaningful questions which the altered theory may be asked to answer.

In the "true" theory, the quantization prescriptions are altered in such a way as to be compatible with a change of all weak relations between true variables into strong equations. Untrue variables which do not even weakly vanish are eliminated; weakly vanishing untrue variables are treated as strongly zero. The theory then can be formulated without ever even mentioning the untrue variables. In Lorentz-covariant quantum electrodynamics, this true theory was first given by Pauli[46] for electromagnetic fields having individual Dirac electrons as source. A generalization of this theory using the quantized (anticommutative) electron field was later developed by Belinfante and Lomont[47] and is known as the "gauge-independent" theory. In this gauge-independent theory, only the transverse components of the electromagnetic field are quantized as a photon field, while the longitudinal part of the electric field is the Coulomb field $-\nabla \int (\rho'/r) d^3\mathbf{x}'$, in which ρ is expressed by (12.23) in terms of the quantized fields of the particles which are its source. The covariance of such a theory is then shown without ever mentioning such "untrue" concepts as that of "longitudina photons."[46,47,42]

Now returning to altered theories, we may ask how weak equations are to be interpreted. Here, we prefer to interpret them as restrictions on Hilbert space,[31,25,41,35] thus avoiding not only non-normalizable state vectors,[31,48] but also the necessity of defining an indefinite metric in Hilbert space.[49] The main advantages of this interpretation are that it makes it easier to justify our frequent use of the assumption (4.6), and that it leads more naturally to the true theory of Sec. 24.

Historically, the existence of the altered theory of quantum electrodynamics was helpful in the development of the true, gauge-independent theory, not only because it led naturally to the elimination of the nonvanishing untrue variables,[45] but also because its formulas[31] were suggestive of formulas needed in the true theory.[47] As a tool in the search for the true theory,

---

[44] D. I. Caplan, Ph.D. thesis, Purdue University (1957).

[45] E. Fermi (1929, 1930). For a treatise in English, see Revs. Modern Phys. 4, 87 (1932). For a treatment avoiding explicit Fourier expansion of the field, see F. J. Belinfante, Physica 12, 1 (1946).

[46] W. Pauli, *Handbuch der Physik* by H. Geiger and H. Scheel (Verlag Julius Springer, Berlin, 1933), second edition, Vol. 24, Part I, Chap. 2, Sec. B8, p. 269. (Reprinted by Edwards Brothers, Inc., Ann Arbor.)

[47] F. J. Belinfante and J. S. Lomont, Phys. Rev. 84, 541 (1951).

[48] F. J. Belinfante, Phys. Rev. 76, 226 (1949).

[49] S. N. Gupta, Proc. Phys. Soc. (London) A63, 681 (1950); A66, 129 (1953).

the altered theory is better *without* than *with* Gupta's interpretation of weak equations.[49] On the other hand, the analog of Gupta's indefinite-metric method may be useful in direct applications of the altered theory as such to physical problems, without attempts at using a "true" quantization procedure.

PART II: THE ALTERED THEORY

15. AUXILIARY CONDITIONS

We now develop an explicit form of the altered theory outlined in the preceding section. Our first task is to choose explicit forms for the auxiliary conditions (14.1); our second task, to choose the K_{mn} of (14.2). Our guiding principles are: (1) We have to destroy by these conditions the complete invariance of our Lagrangian, but like to retain some limited invariance, under a more restricted group of transformations. (2) We like to keep L_{add} simple. (3) Equations (14.6) or (3.1) should be simple enough so that we can solve them without too much trouble for the field velocities. (4) The relation between the constraints and the auxiliary conditions may be kept fairly simple. With these principles of convenience in mind, we have made the following choices.

As a gauge condition, one would like to use the Lorentz condition,

$$0 = \mathfrak{A}^{\nu}{}_{,\nu} \equiv (\mathfrak{g}^{\mu\nu}A_{\mu})_{,\nu}. \tag{15.1}$$

If we would use the square of this in L_{add}, we would be introducing products of $\mathfrak{g}^{\mu\nu}{}_{,\nu}$ and of $A_{\alpha,\beta}$. Consequently, by (14.6), the Maxwell momenta P^{α} would depend on the gravity velocities, and the gravity momenta $p_{\mu}{}^{(\alpha)}$ would depend on the Maxwell velocities (time derivatives of the Maxwell field). This would unnecessarily complicate the solution for the field velocities in terms of the momenta. We can avoid this complication by taking for our gauge condition

$$0 = S \equiv \mathfrak{g}^{\mu\nu}A_{\mu,\nu}, \tag{15.2}$$

and taking for our coordinate condition De Donder's[43]

$$0 = S^{\mu} \equiv \mathfrak{g}^{\mu\nu}{}_{,\nu}. \tag{15.3}$$

Then, the Lorentz condition remains valid by $0 = (S + S^{\mu}A_{\mu})$; but, when L_{add} is made quadratic in S and quadratic in S^{μ}, we avoid cross terms between derivatives of the gravitational field and of the Maxwell field. Moreover, we have thus chosen a coordinate condition beloved by many authors.[50–52,2]

A convenient form for the six fourleg conditions which are to replace the six fourleg constraints (3.10) is

$$0 = S_{(\alpha)(\beta)} \equiv N^{\lambda}[h_{(\alpha)\mu}h_{(\beta)}{}^{\mu}{}_{,\lambda} - h_{(\beta)\mu}h_{(\alpha)}{}^{\mu}{}_{,\lambda}]. \tag{15.4}$$

Here, the N^{λ} are unit vectors tangential to some chosen set of time-like curves of which one runs through every point of space time. They may be normalized by

$$N_{\nu}N^{\nu} \equiv g_{\mu\nu}N^{\mu}N^{\nu} \equiv -1, \tag{15.5a}$$

$$N_0 < 0 < N^0. \tag{15.5b}$$

In order to understand the geometrical meaning of the fourleg conditions (15.4), we multiply them by $h^{(\beta)}{}_{\nu}$ and we use the identity

$$h_{(\alpha)\mu}h_{(\beta)}{}^{\mu}{}_{,\lambda}h^{(\beta)}{}_{\nu} \equiv -h_{(\alpha)\mu,\lambda}h_{(\beta)}{}^{\mu}h^{(\beta)}{}_{\nu} \equiv -h_{(\alpha)\nu,\lambda}$$
$$\equiv (g_{\mu\nu}h_{(\alpha)}{}^{\mu})_{,\lambda}. \tag{15.6}$$

Multiplying the resulting expression by $-\frac{1}{2}g^{\rho\nu}$, we get

$$0 = N^{\lambda}\{\tfrac{1}{2}g^{\rho\nu}g_{\mu\nu,\lambda}h_{(\alpha)}{}^{\mu} + h_{(\alpha)}{}^{\rho}{}_{,\lambda}\}, \tag{15.7}$$

so, by (3.4b) with (3.4c),

$$N^{\lambda}h_{(\alpha)}{}^{\rho}{}_{,\lambda} = N^{\lambda}h_{(\alpha)}{}^{\mu}(-\{\lambda\mu,\rho\}$$
$$+\tfrac{1}{2}g^{\rho\nu}[g_{\lambda\nu,\mu} - g_{\lambda\mu,\nu}]). \tag{15.8}$$

By $g^{\rho\nu} \equiv h^{(\beta)\rho}h_{(\beta)}{}^{\nu}$, this may be written as

$$N^{\lambda}h_{(\alpha)}{}^{\rho}{}_{,\lambda} = -N^{\lambda}\{\lambda\mu,\rho\}h_{(\alpha)}{}^{\mu} + \epsilon_{(\alpha)(\beta)}h^{(\beta)\rho}, \tag{15.9}$$

where we put

$$\epsilon_{(\alpha)(\beta)} \equiv \tfrac{1}{2}[h_{(\alpha)}{}^{\mu}\,h_{(\beta)}{}^{\nu} - h_{(\alpha)}{}^{\nu}\,h_{(\beta)}{}^{\mu}]g_{\lambda\nu,\mu}N^{\lambda}. \tag{15.10}$$

Not only does (15.9) follow from (15.4), but (15.4) also follows from (15.9). This may be seen by multiplying (15.9) by $h_{(\beta)\rho}$, antisymmetrizing the result with respect to (α) and (β), and using (15.10), (3.4b), and

$$[\lambda\nu,\mu] - [\lambda\mu,\nu] = g_{\lambda\mu,\nu} - g_{\lambda\nu,\mu}. \tag{15.11}$$

Therefore, (15.9) is fully equivalent to the fourleg condition (15.4).

From (15.9) we see the meaning of the fourleg condition. It predicts the fourleg field at later times, once the fourleg field has been given at an initial time. It imposes the condition that the fourlegs, at the various points along any one of the time-like curves of direction N^{μ}, are obtained from each other by parallel displacement combined with an additional twist which, per unit of proper time along these curves, is given by the $\epsilon_{(\alpha)(\beta)}$.

This twist, however, is not invariant, but depends on the world coordinate system being used. That is, under a mere coordinate transformation with

$$\partial x^{\mu'}/\partial x^{\nu} = A^{\mu'}{}_{\nu}, \quad \partial^2 x^{\mu'}/\partial x^{\rho}\partial x^{\sigma} = A^{\mu'}{}_{\rho\sigma},$$
$$\partial x^{\sigma}/\partial x^{\nu'} = A^{\sigma}{}_{\nu'}, \quad \text{etc.}, \tag{15.12}$$

and without any change of the leg fields $h_{(\alpha)}$ or of the time-like field N except the coordinate transformation of their components, the twists $\epsilon_{(\alpha)(\beta)}$ transform into

$$\epsilon'_{(\alpha)(\beta)} = \epsilon_{(\alpha)(\beta)} + \tfrac{1}{2}h_{(\alpha)}{}^{\nu'}\,h_{(\beta)}{}^{\mu'}\,N^{\lambda'}g_{\rho\sigma}$$
$$\times(A^{\sigma}{}_{\mu'}A^{\rho}{}_{\nu'\lambda'} - A^{\sigma}{}_{\nu'}A^{\rho}{}_{\mu'\lambda'}). \tag{15.13}$$

The twists, therefore, may alter under nonaffine coordinate transformations, even if by

$$A^{\mu'}{}_{\mu\nu}\mathfrak{g}^{\mu\nu} = 0 \tag{15.14}$$

[50] V. Fock, J. Phys. (USSR) 1, 81 (1939).
[51] A. Papapetrou, Proc. Roy. Irish Acad. A52, 11 (1948).
[52] S. N. Gupta, Proc. Phys. Soc. (London) A65, 608 and 161 (1952).

one keeps the coordinate condition (15.3) valid after as well as before the coordinate transformation. Therefore, the fourleg conditions (15.4) make the fourleg field at later times depend not only on the choice of time-like curves N^λ and on the choice of the initial fourleg field, but also on the choice of the world coordinate system in space time.

One could have avoided the noninvariant twist $\epsilon_{(\alpha)(\beta)}$ in the fourleg field, by choosing, instead of (15.4), the auxiliary conditions

$$0 = N^\lambda[h_{(\alpha)\mu}h_{(\beta)}{}^{\mu}{}_{,\lambda} - h_{(\beta)\mu}h_{(\alpha)}{}^{\mu}{}_{,\lambda}] + N^\lambda g_{\lambda\mu,\nu}[h_{(\alpha)}{}^{\mu}h_{(\beta)}{}^{\nu} - h_{(\beta)}{}^{\mu}h_{(\alpha)}{}^{\nu}]. \quad (15.15)$$

However, use of this expression would complicate L_{add} in such a manner that it would no longer be simple to solve the (3.1) for the $h_{(\beta)}{}^{\nu}{}_{,0}$ in terms of the $p_\mu{}^{(\alpha)}$. We therefore content ourselves with the use of (15.4).

16. RESTRICTED INVARIANCE TRANSFORMATIONS

Our altered Lagrangian $\int Ld^4x$ will be weakly invariant (but for boundary effects) under those of the transformations (3.7), (3.9), (3.11)–(3.12) or combinations thereof, which leave the auxiliary conditions invariant. We calculate the combined effect of these transformations on the expressions (15.2)–(15.4) for the S^n. First, we calculate

$$\bar{\delta}\mathfrak{g}^{\mu\nu} \equiv \xi^{\mu}{}_{,\lambda}\mathfrak{g}^{\lambda\nu} + \xi^{\nu}{}_{,\lambda}\mathfrak{g}^{\lambda\mu} - \xi^{\lambda}{}_{,\lambda}\mathfrak{g}^{\mu\nu} - \xi^{\lambda}\mathfrak{g}^{\mu\nu}{}_{,\lambda}. \quad (16.1)$$

Then, we use the fact that for any quantity F

$$\bar{\delta}(F_{,\nu}) \equiv (\bar{\delta}F)_{,\nu}, \quad (16.2)$$

if $\bar{\delta}$, as in (3.12), denotes the "substantial variation"

$$\bar{\delta}F \equiv \delta F - \xi^\nu F_{,\nu}, \quad \text{with} \quad \xi^\nu = \delta x^\nu. \quad (16.3\text{a–b})$$

Thus we find

$$\bar{\delta}S \equiv -\xi^\lambda S_{,\lambda} - \xi^{\lambda}{}_{,\lambda}S - \xi^{\lambda}{}_{,\mu\nu}\mathfrak{g}^{\mu\nu}A_\lambda + \mathfrak{g}^{\mu\nu}\xi_{,\mu\nu}, \quad (16.4\text{a})$$

$$\bar{\delta}S^\mu \equiv -\xi^\lambda S^{\mu}{}_{,\lambda} - \xi^{\lambda}{}_{,\lambda}S^\mu + \xi^{\mu}{}_{,\lambda}S^\lambda + \mathfrak{g}^{\lambda\nu}\xi^{\mu}{}_{,\lambda\nu}, \quad (16.4\text{b})$$

$$\begin{aligned}\bar{\delta}S_{(\alpha)(\beta)} \equiv &-\xi^\lambda S_{(\alpha)(\beta),\lambda} - \xi_{(\alpha)(\gamma)}\mathring{g}^{(\gamma)(\delta)}S_{(\beta)(\delta)} \\ &+ \xi_{(\beta)(\gamma)}\mathring{g}^{(\gamma)(\delta)}S_{(\alpha)(\delta)} \\ &+ N^\lambda\xi^{\mu}{}_{,\lambda\nu}[h_{(\alpha)\mu}h_{(\beta)}{}^{\nu} - h_{(\beta)\mu}h_{(\alpha)}{}^{\nu}] \\ &- 2N^\lambda\xi_{(\alpha)(\beta),\lambda}. \quad (16.4\text{c})\end{aligned}$$

Therefore, if the auxiliary conditions (14.1) are to be valid after as well as before these infinitesimal transformations, the descriptors ξ, ξ^μ, and $\xi_{(\alpha)(\beta)}$ will have to be restricted by the conditions

$$g^{\mu\nu}\xi_{,\mu\nu} = 0, \qquad g^{\lambda\nu}\xi^{\mu}{}_{,\lambda\nu} = 0, \quad (16.5\text{a–b})$$

and

$$N^\lambda\xi_{(\alpha)(\beta),\lambda} = \tfrac{1}{2}N^\lambda\xi^{\mu}{}_{,\lambda\nu}[h_{(\alpha)\mu}h_{(\beta)}{}^{\nu} - h_{(\beta)\mu}h_{(\alpha)}{}^{\nu}]. \quad (16.5\text{c})$$

The conditions (16.5b) were already mentioned in (15.14). Equations (16.5c) express the necessity of accompanying nonaffine coordinate transformations ($\xi^{\mu}{}_{,\lambda\nu} \neq 0$) by a change in the "time-dependence" of the fourleg field along the time-like curves in the N^λ direction. This necessity stems from the lack of coordinate-invariance of the fourleg condition (15.4), which was also expressed by the property (15.13) of the $\epsilon_{(\alpha)(\beta)}$ occurring in (15.9).

17. ALTERED LAGRANGIAN

We now choose the coefficients K_{mn} of (14.2). We do not want to cause more lack of invariance of the Lagrangian than purposely already introduced by our definition of the auxiliary conditions. The obvious choice for L_{add} is

$$L_{\text{add}} \equiv L_c + L_4 + L_L, \quad (17.1)$$

where the coordinate-condition terms L_c, the fourleg-condition terms L_4, and the Lorentz-condition terms L_L are given by

$$L_c \equiv -\tfrac{1}{2}C\, j^{-1}\, g_{\mu\nu}S^\mu S^\nu, \quad (17.1\text{a})$$

$$L_4 \equiv -\tfrac{1}{4}C\, j\, S_{(\alpha)(\beta)}S^{(\alpha)(\beta)}, \quad (17.1\text{b})$$

$$L_L \equiv -(8\pi j)^{-1}\, SS, \quad (17.1\text{c})$$

with C given by (8.4). Here, the powers of $j \equiv (-g)^{\frac{1}{2}}$ have been chosen in such a way that, including the factors j hidden in $\mathfrak{g}^{\mu\nu}$ in the definitions (15.2)–(15.3) of the S and the S^μ, each term of L is linear in j. The constant coefficients in L_L and in L_c have been given such values that in first approximation the flat-space Lagrangians of Fermi's quantum electrodynamics[45] and of Gupta's gravitational theory[52] are obtained, with our own normalization of the variables. The constant in L_4 was taken in such a manner that there is some cancellation of terms in the special case that N^μ is normal to the hypersurface $x^0 =$ constant, i.e., if

$$N_s = 0, \quad (17.2\text{a})$$

$$N_0 = -(-g^{00})^{-\frac{1}{2}}, \quad (17.2\text{b})$$

$$N^0 = +(-g^{00})^{\frac{1}{2}}. \quad (17.2\text{c})$$

Inserting (15.2)–(15.4) in (17.1a), (17.1b), and (17.1c), and using (3.3)–(3.5) and

$$g_{\lambda\mu,\nu} \equiv -(g_{\lambda\sigma}h^{(\alpha)}{}_{\mu} + g_{\mu\sigma}h^{(\alpha)}{}_{\lambda})h_{(\alpha)}{}^{\sigma}{}_{,\nu}, \quad (17.3\text{a})$$

$$\mathfrak{g}^{\lambda\mu}{}_{,\nu} \equiv j\{h^{(\alpha)\lambda}\,\delta_\sigma{}^\mu + h^{(\alpha)\mu}\,\delta_\sigma{}^\lambda - g^{\lambda\mu}\,h^{(\alpha)}{}_{\sigma}\}h_{(\alpha)}{}^{\sigma}{}_{,\nu}, \quad (17.3\text{b})$$

we find

$$\begin{aligned}L_g + L_c \equiv\ &\tfrac{1}{4}Cj\, g_{\alpha\beta,\nu}g_{\rho\sigma,\tau}[2g^{\alpha\tau}g^{\beta\sigma}g^{\nu\rho} - g^{\alpha\rho}g^{\beta\sigma}g^{\nu\tau} \\ &+\tfrac{1}{2}g^{\alpha\beta}g^{\rho\sigma}g^{\nu\tau} - 2g^{\alpha\nu}g^{\beta\rho}g^{\sigma\tau}] \\ \equiv\ &\tfrac{1}{2}Cj\, h_{(\alpha)}{}^{\mu}{}_{,\sigma}\, h_{(\beta)}{}^{\nu}{}_{,\tau}\,[g_{\mu\nu}\, h^{(\alpha)\tau}\, h^{(\beta)\sigma} \\ &- g_{\mu\nu}\, h^{(\alpha)\sigma}\, h^{(\beta)\tau} + g^{\sigma\tau}\, h^{(\alpha)}{}_{\mu}\, h^{(\beta)}{}_{\nu} \\ &- g^{\sigma\tau}\, h^{(\alpha)}{}_{\nu}\, h^{(\beta)}{}_{\mu} - g_{\mu\nu}g^{\sigma\tau}\mathring{g}^{(\alpha)(\beta)} \\ &+ 2h^{(\alpha)}{}_{\nu}\, h^{(\beta)\sigma}\,\delta_\mu{}^\tau - 2h^{(\beta)}{}_{\mu}\, h^{(\alpha)\sigma}\,\delta_\nu{}^\tau \\ &+\mathring{g}^{(\alpha)(\beta)}\,\delta_\nu{}^\sigma\,\delta_\mu{}^\tau - \mathring{g}^{(\alpha)(\beta)}\,\delta_\mu{}^\sigma\,\delta_\nu{}^\tau], \quad (17.4\text{a})\end{aligned}$$

$$L_4 \equiv \tfrac{1}{2}Cj\, h_{(\alpha)}{}^{\mu}{}_{,\sigma}\, h_{(\beta)}{}^{\nu}{}_{,\tau}\, N^\sigma N^\tau\,[h^{(\alpha)}{}_{\nu}\, h^{(\beta)}{}_{\mu} - \mathring{g}^{(\alpha)(\beta)}g_{\mu\nu}], \quad (17.4\text{b})$$

$$L_f + L_L \equiv (8\pi)^{-1} j A_{\alpha,\beta} A_{\mu,\nu} [g^{\alpha\nu} g^{\beta\mu} - g^{\alpha\mu} g^{\beta\nu} - g^{\alpha\beta} g^{\mu\nu}]. \quad (17.4c)$$

The altered Lagrangian density then is the sum of (17.4a), (17.4b), (17.4c), (3.3d), (3.5b), and the last term of (12.15). Here, (3.4h) and (12.12a–c) with (12.9)–(12.10) should be used for expressing $\tilde{L}$ in terms of the $\tilde{q}$ and $\tilde{q}_{,\mu}$.

18. $\dot{q}(p)$ OF THE ALTERED THEORY

The altered theory enables us to express the field velocities $\tilde{q}_{,0}$ in terms of the canonical variables $\tilde{q}$ and $\tilde{p}$. The equation for $\tilde{\psi}_{,0}$ is obtained by variation of $\tilde{\psi}^\dagger$ in $\tilde{L}$ as given at the end of the preceding section.

For the electromagnetic field, the $\tilde{q}$ and $\tilde{p}$ are identical with the q and p. We find for the canonical field momenta

$$P^\mu \equiv (4\pi)^{-1} j A_{\alpha,\beta} [g^{\alpha 0} g^{\beta\mu} - g^{\beta 0} g^{\alpha\mu} - g^{\mu 0} g^{\alpha\beta}]. \quad (18.1)$$

These equations are solved by

$$A_{\alpha,0} \equiv -4\pi (\mathfrak{g}^{00})^{-1} g_{\alpha\mu} P^\mu + (g^{00})^{-1} A_{\beta,s} [\delta_\alpha{}^s g^{\beta 0} - \delta_\alpha{}^\beta g^{s0} - \delta_\alpha{}^0 g^{\beta s}]. \quad (18.2)$$

For the gravitational field (fourleg field), we first calculate the old field momenta

$$\begin{aligned} p_\mu{}^{(\alpha)} \equiv C j\, h_{(\beta)}{}^\nu{}_{,\tau} [& g_{\mu\nu} h^{(\alpha)\tau} h^{(\beta)0} - g_{\mu\nu} h^{(\alpha)0} h^{(\beta)\tau} \\ & + g^{\tau 0} h^{(\alpha)}{}_\mu h^{(\beta)}{}_\nu - g^{\tau 0} h^{(\alpha)}{}_\nu h^{(\beta)}{}_\mu \\ & - g_{\mu\nu} g^{\tau 0} \mathring{g}^{(\alpha)(\beta)} + h^{(\alpha)}{}_\nu h^{(\beta)0} \delta_\mu{}^\tau + h^{(\beta)}{}_\mu h^{(\alpha)\tau} \delta_\nu{}^0 \\ & - h^{(\beta)}{}_\mu h^{(\alpha)0} \delta_\nu{}^\tau - h^{(\alpha)}{}_\nu h^{(\beta)\tau} \delta_\mu{}^0 + \mathring{g}^{(\alpha)(\beta)} \delta_\mu{}^\tau \delta_\nu{}^0 \\ & - \mathring{g}^{(\alpha)(\beta)} \delta_\nu{}^\tau \delta_\mu{}^0 + h^{(\alpha)}{}_\nu h^{(\beta)}{}_\mu N^\tau N^0 \\ & - \mathring{g}^{(\alpha)(\beta)} g_{\mu\nu} N^\tau N^0] \\ & + \tfrac{1}{8} \hbar c j \bar{\psi} (\gamma_\mu \gamma^0 \gamma^{(\alpha)} - \gamma^{(\alpha)} \gamma^0 \gamma_\mu) \psi. \end{aligned} \quad (18.3)$$

These 16 equations for the $\tilde{q}_{,0} \equiv h_{(\beta)}{}^\nu{}_{,0}$ are solved by[44]

$$\begin{aligned} h_{(\beta)}{}^\nu{}_{,0} \equiv & (4Cj)^{-1} p_\mu{}^{(\alpha)} [(g^{00})^{-1} (h_{(\alpha)}{}^\mu h_{(\beta)}{}^\nu \\ & - h_{(\beta)}{}^\mu h_{(\alpha)}{}^\nu - g^{\mu\nu} \mathring{g}_{(\alpha)(\beta)}) \\ & + (N^0)^{-2} (h_{(\beta)}{}^\mu h_{(\alpha)}{}^\nu - g^{\mu\nu} \mathring{g}_{(\alpha)(\beta)})] \\ & + \tfrac{1}{2} h_{(\alpha)}{}^\mu{}_{,s} [(N^s/N^0)(h^{(\alpha)\nu} h_{(\beta)\mu} \\ & - \delta_{(\beta)}{}^{(\alpha)} \delta_\mu{}^\nu) + (g^{00})^{-1} (h^{(\alpha)0} h_{(\beta)\mu} g^{\nu s} \\ & + h^{(\alpha)0} h_{(\beta)}{}^s \delta_\mu{}^\nu - h^{(\alpha)s} h_{(\beta)\mu} g^{\nu 0} \\ & - h^{(\alpha)s} h_{(\beta)}{}^0 \delta_\mu{}^\nu - \delta_{(\beta)}{}^{(\alpha)} \delta_\mu{}^\nu g^{s0} \\ & - \delta_{(\beta)}{}^{(\alpha)} \delta_\mu{}^s g^{\nu 0} + \delta_{(\beta)}{}^{(\alpha)} \delta_\mu{}^0 g^{\nu s} \\ & - h^{(\alpha)\nu} h_{(\beta)\mu} g^{s0} - h^{(\alpha)\nu} h_{(\beta)}{}^0 \delta_\mu{}^s \\ & + h^{(\alpha)\nu} h_{(\beta)}{}^s \delta_\mu{}^0)] \\ & + \pi \hbar G c^{-3} (N^0)^{-2} \bar{\psi} (\gamma^\nu \gamma^0 \gamma_{(\beta)} - \gamma_{(\beta)} \gamma^0 \gamma^\nu) \psi. \end{aligned} \quad (18.4)$$

This can be expressed in terms of the $\tilde{q}$ and $\tilde{p}$ by means of (12.12a–c) and (12.20). Thus we have obtained all field velocities in terms of the canonical variables.

At this point, a few remarks may be made about the so-called "Feynman quantization" of our theory.[53,54] In this formalism, integrations are performed over field variables. According to B. Laurent,[55] the integrations over the gravitational field should be interpreted as integrations over all possible values of the sixteen fourleg components $h_{(\alpha)}{}^\mu$. According to B. S. DeWitt[55] and to J. L. Anderson,[55] the "volume element in q space" should be a product of the differentials of the q and a weight factor depending on the determinant with elements

$$G^{ij} \equiv \partial^2 L / \partial q_{i,0} \partial q_{j,0}. \quad (18.5)$$

If for L one chooses the basic Lagrangian $L_\circ$, then one meets the difficulty that the determinant of the G^{ij} vanishes. In the altered theory, however, this difficulty is avoided. The determinant then can be factorized into a determinant G_h where the q_i are the $h_{(\alpha)}{}^\mu$, a determinant G_A where the q_i are the A_μ, and a determinant G_e for the electron field. We find

$$G_h \equiv C^{16} (\mathfrak{g}^{00})^{16} \cdot \mathrm{Det}\{ (1+n) h^{(\alpha)}{}_\nu h^{(\beta)}{}_\mu - h^{(\alpha)}{}_\mu h^{(\beta)}{}_\nu + (1-n) g_{\mu\nu} \mathring{g}^{(\alpha)(\beta)} \}, \quad (18.6)$$

where the rows of the determinant are labeled by (α) and μ and the columns by (β) and ν, and where

$$n \equiv (N^0)^2 (-g^{00})^{-1}. \quad (18.7)$$

The fact that this determinant remains a function of the field variables $h_{(\alpha)}{}^\mu$ themselves may cause difficulties in attempts at verifying whether the procedure of Feynman quantization of the altered theory of gravitation will give results equivalent with those which one would obtain from a similar procedure using merely two true variables instead of the sixteen variables of the altered theory. In Fermi's quantum electrodynamics, this complication is avoided because

$$G_A \equiv (4\pi)^{-4} (g^{00})^4 g \quad (18.8)$$

does not depend on the A_μ themselves, thus enabling Laurent[55] to show the equivalence of the two corresponding procedures in quantum electrodynamics.

19. ANALOGY OF CONSTRAINTS AND AUXILIARY CONDITIONS

In Sec. 14 we have shown that in the altered theory the constraints of Part I hold weakly on account of the auxiliary conditions (14.1). We now show that, conversely, the auxiliary conditions (14.1) will be satisfied whenever the primary constraints

$$\Phi^k = 0 \quad (19.1)$$

are given to hold weakly. Here, the Φ^k are given by (6.8a), (6.8b), and (6.8c).

For proving this statement we substitute, in the definitions (15.2)–(15.4) of the quantities S^n, (18.2) and (18.4) for $A_{\alpha,0}$ and for $h_{(\beta)}{}^\nu{}_{,0}$, after first expressing

---

[53] R. P. Feynman, Phys. Rev. 76, 769 (1949); 80, 440 (1950).

[54] W. Pauli, *Ausgewählte Kapitel aus der Feldquantisierung* (Verlagsauslieferung Akademische Buchgenossenschaft, Zürich, 1951).

[55] Oral communication, Chapel Hill Conference on the Role of Gravitation in Physics, January 18–23, 1957.

derivatives of the metrical tensor by (17.3). Thus we find

$$S \equiv -4\pi(g^{00})^{-1}P^0 \equiv -4\pi(g^{00})^{-1}\Phi, \qquad (19.2a\text{–}b)$$

$$S^\mu \equiv -(2C)^{-1}(g^{00})^{-1}p_\lambda{}^{(\alpha)}[g^{\lambda 0}\, h_{(\alpha)}{}^\mu + g^{\lambda\mu}\, h_{(\alpha)}{}^0] + (g^{00})^{-1}[g^{\mu s}\, g^{00}{}_{,s} - g^{0\mu}\, g^{0s}{}_{,s}] \qquad (19.3a)$$

$$\equiv (2Cg^{00})^{-1}[h_{(\alpha)}{}^0\, h_{(\beta)}{}^\mu\, \Phi^{(\alpha)(\beta)} - 2g^{\mu\lambda}\Phi_\lambda], \qquad (19.3b)$$

$$S_{(\alpha)(\beta)} \equiv (2CN^0 j)^{-1}\, p_\mu{}^{(\gamma)}[\mathring{g}_{(\alpha)(\gamma)}h_{(\beta)}{}^\mu - \mathring{g}_{(\beta)(\gamma)}h_{(\alpha)}{}^\mu] + 2\pi\hbar G c^{-3}(N^0)^{-1}\,\bar{\psi}[\gamma_{(\alpha)}\gamma^0\gamma_{(\beta)} - \gamma_{(\beta)}\gamma^0\gamma_{(\alpha)}]\psi \qquad (19.4a)$$

$$\equiv (2CN^0 j)^{-1}\, \mathring{g}_{(\alpha)(\gamma)}\mathring{g}_{(\beta)(\delta)}\Phi^{(\gamma)(\delta)}. \qquad (19.4b)$$

This proves that the auxiliary conditions of the altered theory are completely equivalent to the constraints of the original theory. Similarly, the time derivatives of the auxiliary conditions are equivalent to combinations of the χ constraints with each other and with the Φ constraints.

20. CONTINUED VALIDITY OF THE AUXILIARY CONDITIONS

A simple proof of the compatibility of the auxiliary conditions with the field equations of motion of the altered theory can be given using the general invariance properties of the basic part $L_\circ$ of the Lagrangian density discussed in Sec. 3.[44] Bergmann[21,22,25] has proved the following theorem. Consider infinitesimal transformations of the coordinates x^μ and the field variables y_A (which is Bergmann's notation for our q_i). At the point P, let these transformations be given by

$$x'^\mu(P) \equiv x^\mu(P) + \delta x^\mu \equiv x^\mu(P) + \xi^\mu(P); \qquad (20.1)$$

$$y_A'(P) \equiv y_A(P) + \delta y_A, \qquad (20.2a)$$

$$\delta y_A \equiv c_{A\mathrm{k}}{}^\nu\, \xi^{\mathrm{k}}{}_{,\nu} + c_{A\mathrm{k}}\xi^{\mathrm{k}}, \qquad (20.2b)$$

where the *descriptors* ξ^{k} with $\mathrm{k}=0, 1, 2, 3, 4, 5, \cdots, 10$ label the transformations. In particular, the descriptors for $\mathrm{k}=\mu(=0, 1, 2, 3)$ label the coordinate transformations (20.1); ξ^4 [also called ξ, see (3.7)] labels the gauge transformations, and ξ^5 through ξ^{10} [also called $\xi_{(\alpha)(\beta)}$, see (3.9)] label the fourleg rotations in the local Lorentz frames. Define the point $\bar{P}$ by

$$x'^\mu(\bar{P}) \equiv x^\mu(P), \qquad (20.3a)$$

so

$$x^\mu(\bar{P}) \equiv x^\mu(P) - \xi^\mu(P), \qquad (20.3b)$$

and put, as in (3.12) and (16.3),

$$y_A'(\bar{P}) \equiv y_A(P) + \bar{\delta} y_A, \qquad (20.4a)$$

$$\bar{\delta} y_A \equiv \delta y_A - y_{A,\mu}\xi^\mu. \qquad (20.4b)$$

Now suppose that, under such general transformations *not* restricted by (16.5), the part L_* of the Lagrangian density at P transforms into

$$L_*'(\bar{P}) \equiv L_*(P) + \bar{\delta} L_* \qquad (20.5a)$$

at $\bar{P}$, and suppose that $\bar{\delta}L_*$ is the four divergence of some four-component quantity Q^ν which need not be a four-vector:

$$\bar{\delta} L_* \equiv Q^\nu{}_{,\nu}. \qquad (20.5b)$$

If L_* has the invariance property (20.5b), then Bergmann proves that automatically

$$L_*{}^A[c_{A\mathrm{k}} - y_{A,\mathrm{k}}] - (L_*{}^A\, c_{A\mathrm{k}}{}^\nu)_{,\nu} \equiv 0, \qquad (20.6)$$

where

$$L_*{}^A \equiv \partial^P L_*/\partial y_A - (\partial^P L_*/\partial y_{A,\nu})_{,\nu}, \qquad (20.7)$$

and where $y_{A,\mathrm{k}}$ for $\mathrm{k}=\nu$ means $y_{A,\nu}$, and for $\mathrm{k}=4, 5, 6, \cdots, 10$ means zero.

We first apply this theorem to the infinitesimal gauge transformations of the original variables $y_A \equiv q_i$, in (3.7). For these transformations we find the $c_{A4}{}^\nu$ and c_{A4} by comparing (3.7) with (20.2b). Under such transformations, (20.5b) is valid, with $\bar{P} \equiv P$ and with $Q^\nu \equiv 0$, not only for $L_\circ$ but also for L_c and L_4 given by (17.1a–b). Therefore in this case taking

$$L_* \equiv L_\circ + L_c + L_4 \equiv L - L_L, \qquad (20.8)$$

we find that (20.6) now takes the form

$$(ie/\hbar c)[L_*{}^\psi\, \psi - L_*{}^{\psi\dagger\beta}\, \bar{\psi}] - L_*{}^\mu{}_{,\mu} \equiv 0, \qquad (20.9)$$

where $L_*{}^\psi$, $L_*{}^{\psi\dagger\beta}$, and $L_*{}^\mu$ are the expressions (20.7) for $y_A \equiv \psi$, for $y_A \equiv \bar{\psi}$, and for $y_A \equiv A_\mu$, respectively. Since L_L does not depend on either ψ or $\bar{\psi}$, we find

$$L_*{}^\psi \equiv L^\psi \equiv 0, \qquad (20.10a)$$

$$L_*{}^{\psi\dagger\beta} \equiv L^{\psi\dagger\beta} \equiv 0, \qquad (20.10b)$$

where we used the electron field equations. On account of the field equations of the altered theory $L^\mu \equiv 0$ for the electromagnetic field, (20.9) then yields

$$0 \equiv -L_*{}^\mu{}_{,\mu} \equiv L_L{}^\mu{}_{,\mu} \equiv [\partial L_L/\partial A_\mu - (\partial L_L/\partial A_{\mu,\nu})_{,\nu}]_{,\mu}. \qquad (20.11)$$

Inserting (17.1c) with (15.2) in (20.11) we obtain

$$0 \equiv (4\pi)^{-1}(Sg^{\mu\nu})_{,\mu\nu}, \qquad (20.12a)$$

so that

$$S_{,00} \equiv (-g^{00})^{-1}[g^{rs}S_{,rs} + 2g^{r0}S_{,r0} + 2g^{\mu\nu}{}_{,\nu}S_{,\mu} + g^{\mu\nu}{}_{,\mu\nu}S]. \qquad (12.12b)$$

Now assume that, at a given time x^0,

$$S(\mathbf{x},x^0) = 0, \qquad (20.13a)$$

$$S_{,0}(\mathbf{x},x^0) = 0. \qquad (20.13b)$$

Thence, at x^0, also $S_{,s} = 0$, $S_{,rs} = 0$, $S_{,r0} = S_{,0s} = 0$ for $r,s = 1, 2, 3$. From (20.12b), then, at x^0,

$$S_{,00} = 0. \qquad (20.14a)$$

The nth time derivative of (20.12b) shows that, if at x^0 for every $\mathbf{x}$ all time derivatives of S up to the $(n+1)$st vanish weakly, then also the $(n+2)$nd time derivative will vanish. Thence, by complete induction, validity

of (20.13a) and (20.13b) at x^0 yields at x^0 also

$$S_{,00}=0, \quad S_{,000}=0, \quad S_{,0000}=0, \text{ etc.} \tag{20.14b}$$

By Taylor expansion we thus find from (20.12) with (20.13a) and (20.13b)

$$S(x) = 0 \text{ for any } x^0. \tag{20.15}$$

This shows that the *one* "all-time" auxiliary condition (15.2) is equivalent to *two* "initial" conditions (20.13a) and (20.13b) imposed at some starting time x^0, from whereon the auxiliary condition will remain valid on account of the field equations of motion used in (20.10)–(20.11).

As a second application of (20.6), we use it for the infinitesimal fourleg rotations (3.9a), (3.9b), (3.9c), and (3.9d), for which the index k stands for the six combinations $(\alpha)(\beta)$ with $(\alpha)<(\beta)$. The $c_{A\mathrm{k}}{}^{\nu}$ this time all are zero, and the $c_{A\mathrm{k}}$ are,

$$\text{for } y_A \equiv h_{(\gamma)}{}^{\mu}, \quad c_{A\mathrm{k}} \equiv h^{(\beta)\mu}\delta_{(\gamma)}{}^{(\alpha)} - h^{(\alpha)\mu}\delta_{(\gamma)}{}^{(\beta)}, \tag{20.16a}$$

$$\text{for } y_A \equiv \psi, \quad c_{A\mathrm{k}} \equiv \tfrac{1}{4}[\gamma^{(\alpha)}\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^{(\alpha)}]\psi, \tag{20.16b}$$

$$\text{for } y_A \equiv \bar{\psi}, \quad c_{A\mathrm{k}} \equiv \tfrac{1}{4}\bar{\psi}[\gamma^{(\beta)}\gamma^{(\alpha)} - \gamma^{(\alpha)}\gamma^{(\beta)}]; \tag{20.16c}$$

all for $(\alpha) < (\beta)$ only. For $(\alpha) > (\beta)$, $c_{A\mathrm{k}} \equiv 0$.

In absence of gauge and coordinate transformations we find that this time not only $L_{\circ}$ but also L_c and L_L have the invariance property (20.5), with $\bar{P} \equiv P$. Therefore, we now choose

$$L^* \equiv L_{\circ} + L_c + L_L \equiv L - L_4. \tag{20.17}$$

Then, (20.6) yields $0 \equiv L_*{}^A c_{A\mathrm{k}}$, or, by the field equations $L^A \equiv 0$,

$$0 \equiv - L_4{}^A c_{A\mathrm{k}} \equiv c_{A\mathrm{k}}[(\partial^P L_4/\partial y_{A,\nu})_{,\nu} - \partial^P L_4/\partial y_A]. \tag{20.18}$$

By (17.1b) and (20.16a) with $(\alpha) < (\beta)$, this gives

$$0 \equiv -\tfrac{1}{4}C\{h^{(\beta)\mu}\delta_{(\gamma)}{}^{(\alpha)} - h^{(\alpha)\mu}\delta_{(\gamma)}{}^{(\beta)}\}$$
$$\times\left[\left(\frac{\partial(S_{(\delta)(\epsilon)}S^{(\delta)(\epsilon)}j)}{\partial h_{(\gamma)}{}^{\mu}{}_{,\nu}}\right)_{,\nu} - \frac{\partial(S_{(\delta)(\epsilon)}S^{(\delta)(\epsilon)}j)}{\partial h_{(\gamma)}{}^{\mu}}\right]. \tag{20.19}$$

Most of (20.19) contains at least one undifferentiated factor $S_{(\delta)(\epsilon)}$ or $S^{(\delta)(\epsilon)}$; the only exceptions are the terms

$$-\tfrac{1}{2}C[h^{(\beta)\mu}\delta_{(\gamma)}{}^{(\alpha)} - h^{(\alpha)\mu}\delta_{(\gamma)}{}^{(\beta)}]$$
$$\times(\partial S_{(\delta)(\epsilon)}/\partial h_{(\gamma)}{}^{\mu}{}_{,\nu})S^{(\delta)(\epsilon)}{}_{,\nu}j, \tag{20.20}$$

which by (15.4) are found to yield $CjN^{\nu}S^{(\alpha)(\beta)}{}_{,\nu}$. Therefore, (20.19) but for a factor takes the form

$$0 \equiv N^{\nu}S^{(\alpha)(\beta)}{}_{,\nu} + \text{terms with factors } S^{(\delta)(\epsilon)}. \tag{20.21}$$

Therefore, if, at some given initial time x^0,

$$S^{(\alpha)(\beta)}(\mathbf{x},x^0) = 0, \tag{20.22}$$

so that also $S^{(\alpha)(\beta)}{}_{,r} = 0$, then by (20.21) at that x^0 also $N^0S^{(\alpha)(\beta)}{}_{,0} = 0$ and therefore $S^{(\alpha)(\beta)}{}_{,0} = 0$ at that x^0. Using a reasoning similar to that between (20.13) and (20.15) we find

$$S^{(\alpha)(\beta)}(x) = 0 \quad \text{for any } x^0, \tag{20.23}$$

as a consequence of the initial condition (20.22) and the field equations of motion used at (20.18) in the derivation of (20.21).

Finally, we apply (20.6) to the infinitesimal coordinate transformations (3.11)–(3.12). Comparison of (3.11) with (20.2b) shows that the $c_{A\mu}$ this time all vanish, and gives us the values of the various $c_{A\mu}{}^{\nu}$. Under these coordinate transformations, the basic Lagrangian density $L_{\circ}$ has the property (20.5b), so we put

$$L_* \equiv L_{\circ} \equiv L - L_{\mathrm{add}} \equiv L - K_{\mathrm{mn}}S^{\mathrm{m}}S^{\mathrm{n}}, \tag{20.24}$$

[see (14.2)]. According to (17.1) the $K_{\mathrm{mn}} \equiv K_{\mathrm{nm}}$ depend on the y_A but not on the $y_{A,\mu}$. On account of the field equations $L^A \equiv 0$, the identities (20.6) this time by (20.24) may be rewritten as

$$0 \equiv L_{\mathrm{add}}{}^A\, y_{A,\mu} + (L_{\mathrm{add}}{}^A\, c_{A\mu}{}^{\nu})_{,\nu}, \tag{20.25}$$

where, by (20.7),

$$L_{\mathrm{add}}{}^A \equiv (\partial K_{\mathrm{mn}}/\partial y_A)S^{\mathrm{m}}S^{\mathrm{n}} + 2K_{\mathrm{mn}}S^{\mathrm{m}}(\partial S^{\mathrm{n}}/\partial y_A)$$
$$- 2[K_{\mathrm{mn}}(\partial S^{\mathrm{m}}/\partial y_{A,\lambda})]_{,\lambda}S^{\mathrm{n}}$$
$$-2K_{\mathrm{mn}}(\partial S^{\mathrm{m}}/\partial y_{A,\lambda})S^{\mathrm{n}}{}_{,\lambda}. \tag{20.26}$$

Now, let us assume that at some initial x^0 it is given that

$$S^{\mathrm{n}} = 0, \quad S^{\mathrm{n}}{}_{,\lambda} = 0 \quad (\text{at } x^0). \tag{20.27}$$

As this assumption includes $S = 0$, $S_{,0} = 0$, and $S^{(\alpha)(\beta)} = 0$, we infer from our previous results that now $S = S^{(\alpha)(\beta)} = 0$ for all x^0, so that also

$$S_{,00} = S^{(\alpha)(\beta)}{}_{,00} = 0. \tag{20.28}$$

But what about $S^{\mu}{}_{,0}$?

Inserting (20.26) in (20.25), we find that all terms contain at least one factor S^{n} or $S^{\mathrm{n}}{}_{,\nu}$ except the terms

$$- 2c_{A\mu}{}^{\nu}\, K_{\mathrm{mn}}(\partial S^{\mathrm{m}}/\partial y_{A,\lambda})S^{\mathrm{n}}{}_{,\lambda\nu}. \tag{20.29}$$

Even these terms vanish according to (20.27) or (20.28) except possibly the terms with n $= \sigma$ $(=0,1,2,3)$ and $\lambda = \nu = 0$, that is,

$$- 2c_{A\mu}{}^{0}\, K_{\mathrm{m}\sigma}(\partial S^{\mathrm{m}}/\partial y_{A,0})S^{\sigma}{}_{,00}. \tag{20.30}$$

From (17.1) we find that necessarily m $= \rho$ $(=0,1,2,3)$ and that $K_{\rho\sigma} \equiv -\tfrac{1}{2}Cj^{-1}g_{\rho\sigma}$. Using

$$S^{\rho} \equiv j[h^{(\beta)\rho}\,\delta_{\lambda}{}^{\nu} + h^{(\beta)\nu}\,\delta_{\lambda}{}^{\rho} - g^{\rho\nu}\,h^{(\beta)}{}_{\lambda}]h_{(\beta)}{}^{\lambda}{}_{,\nu} \tag{20.31}$$

for calculating $(\partial S^{\rho}/\partial y_{A,0})$ for $y_A \equiv h_{(\beta)}{}^{\lambda}$, for which $c_{A\mu}{}^{\nu} \equiv h_{(\beta)}{}^{\nu}\delta_{\mu}{}^{\lambda}$, we find for (20.30)

$$C\,h_{(\beta)}{}^{0}\,\delta_{\mu}{}^{\lambda}\,g_{\rho\sigma}[h^{(\beta)\rho}\,\delta_{\lambda}{}^{0} + h^{(\beta)0}\delta_{\lambda}{}^{\rho} - g^{\rho 0}\,h^{(\beta)}{}_{\lambda}]S^{\sigma}{}_{,00}$$
$$\equiv C\,g^{00}\,g_{\mu\sigma}\,S^{\sigma}{}_{,00}. \tag{20.32}$$

Equation (20.25) therefore tells us that (20.32) must weakly vanish whenever (20.27) is weakly valid, so that the assumptions (20.27) lead to $S^{\mu}{}_{,00}=0$ at x^0. By an argument similar to that between (20.13) and (20.15) we conclude that

$$S^{\mu}(x)=0 \quad \text{for any } x^0 \tag{20.33}$$

follows from the initial conditions (20.27) at some given x^0. Therefore the eleven all-time conditions (20.15), (20.23), and (20.33) are equivalent to sixteen initial conditions (20.13), (20.22), and (20.27) for $\mathrm{n}=\mu$, since the postulate $S^{(\alpha)(\beta)}{}_{,\lambda}=0$ in (20.27) follows already from (20.22) by (20.23). The equations of motion $L^A=0$ are responsible for this continued validity of the auxiliary conditions.

Throughout these derivations we have used tacitly the assumption (4.6) that products containing a factor S^{n} will vanish weakly, whatever the position of the factor may be. Compare Sec. 14.

21. CANONICAL QUANTIZATION OF THE ALTERED THEORY

We quantize the altered theory by the canonical commutation relations (12.1) between the canonical variables $\tilde{q}$ and $\tilde{p}$, which is equivalent to the quantization of the original q and $p=\partial^P L/\partial q_{,0}$ by means of modified P.B. as in Sec. 10.

As the S^{n} now will not commute with all field variables, the tacit use we have been making of the assumption (4.6) may not always be justified. We shall not investigate this point, as we trust that any errors existing in this respect in the altered theory will automatically be removed as in Sec. 24 we go over to the true theory, in which all terms containing factors S^{n} are systematically omitted from the formalism.

22. COVARIANCE

As in Sec. 13, for a proof of the covariance of the commutation relations it suffices to prove that, on a hypersurface $x^0=$ constant, all time derivatives of the commutation relations will vanish whenever the commutation relations themselves are valid on that hypersurface.

Starting from the Lagrangian function $\tilde{L}$ and using the canonical variables $\tilde{q}$ and $\tilde{p}$ and the Hamiltonian defined by (14.10), we can derive the conventional equations of Hamilton and, thence, using (12.1) at x^0, the Heisenberg equations of motion[10]

$$i\hbar c F_{,0}\equiv[F\,;\,\mathfrak{H}]. \tag{22.1}$$

In our altered theory, formulated in terms of the canonical variables $\tilde{q}$ and $\tilde{p}$, there are no constraints interfering with the derivation of (22.1). Also, any function of the field variables and of any derivatives of them can be expressed in terms of these canonical variables and their spatial derivatives.

If the commutation relations (12.1) hold as strong equations on the hypersurface at x^0, the time derivatives of the commutators, according to (22.1) equal to $(i\hbar c)^{-1}\times$ their commutators with $\mathfrak{H}$, are equal to commutators of c numbers with $\mathfrak{H}$, and therefore vanish on this hypersurface. Applying (22.1) at x^0 to $F=$these time derivatives, we find the vanishing of the second derivatives of the commutation relations, and so on. By Taylor expansion we thus find the validity of the postulated canonical commutation relations and therefore of the Heisenberg equations of motion at all times.

Since any space-like surface can be chosen as a surface $x^0=$ constant (even with the coordinate choice restricted by the De Donder condition), we thus have formally proved the covariance of the commutation relations in the altered theory.

As already mentioned below (14.11), in the altered theory the field equations themselves are not completely covariant. Given two surfaces $x^0=$ constant, the altered Lagrangian between them will depend on the frame of reference, the fourleg field, and the gauge chosen between these two surfaces, since L_{add} was purposely chosen to be not generally invariant under transformations which undo the auxiliary conditions. If, therefore, we integrate the field equations and thus express the field variables on the second surface in terms of those on the first surface, we will get results which are not independent of the choice of coordinates and gauge in between. Whatever the meaning of the variables on the second surface thus turns out to be, in any case we have shown above that they will satisfy again the same (form of) commutation relations. That is, the meaning of the $\tilde{q}$ and $\tilde{p}$ may vary, but not the relations (12.1) between them: *formally*, the relations (12.1) are fully invariant.

Differences in meaning of the field variables, however, are due completely to L_{add} and to the terms resulting from L_{add} in the field equations, which in the altered theory may be given in the form of the Heisenberg equations (22.1). These equations are formally integrated by

$$F(t)\equiv U(t)F(0)U(t)^{-1}, \tag{22.2}$$

$$U(t)\equiv 1-\int_0^t\frac{dt'}{i\hbar}\mathfrak{H}(t')+\int_0^t\frac{dt'}{i\hbar}\mathfrak{H}(t')\times\int_0^{t'}\frac{dt''}{i\hbar}\mathfrak{H}(t'')-\cdots+\cdots \tag{22.3a}$$

$$\equiv(1-\mathfrak{H}_n dt_n/i\hbar)(1-\mathfrak{H}_{n-1}dt_{n-1}/i\hbar)\cdots\times\cdots(1-\mathfrak{H}_2 dt_2/i\hbar)(1-\mathfrak{H}_1 dt_1/i\hbar). \tag{22.3b}$$

Here, dt_1, dt_2, $\cdots$, dt_n (with $n\to\infty$) stand for the infinitesimal time intervals (in chronological order) into which the time t may be thought subdivided; $\mathfrak{H}_n$

stands for $\tilde{\mathcal{H}}(t)$ in the time interval between

$$t=\sum_{i=1}^{n-1} dt_i \quad \text{and} \quad t+dt=\sum_{i=1}^{n} dt_i.$$

If in (22.3) we would replace $\tilde{\mathcal{H}}(t)$ by $\tilde{\mathcal{H}}_\circ(t)$, so that consequently $U(t)$ would be changed into a quantity $U_\circ(t)$, then $F_\circ(t) \equiv U_\circ(t)F(0)U_\circ(t)^{-1}$ would satisfy

$$i\hbar c F_{\circ,0}(t) \equiv [F_\circ(t)\ ;\ \tilde{\mathcal{H}}_\circ(t)], \qquad (22.4)$$

so, in the classical limit, in which $\tilde{\mathcal{H}}_\circ$ is distinguished from $\mathcal{H}_\circ$ merely by its being expressed in terms of different variables (see Sec. 12), the prescription (9.3) would yield

$$F_{\circ,0}(t) \equiv [F_\circ(t),\tilde{\mathcal{H}}_\circ(t)] \equiv [F_\circ(t),\mathcal{H}_\circ(t)] \quad \text{for } \hbar\rightarrow 0. \qquad (22.5)$$

Except for terms with o_i, this is the time dependence (9.15) of $F(t)$ of the covariant (unaltered) theory.

The unaltered theory does not predict the time dependence of all variables. As explained in (9.18), the time dependence of quantities like (for instance) A_0 is left undetermined in the unaltered theory in absence of auxiliary conditions. Such quantities, of which this unaltered theory cannot predict the future, must be considered physically meaningless (unobservable), if the unaltered theory is believed to describe nature. There are, however, some other quantities (compare the true variables discussed in Sec. 24) of which the unaltered field equations *do* describe the future in an unambiguous way.

If we now ask in what ways the predictions (22.2) by our altered theory, for the latter meaningful quantities, differ from the predictions made by the original, unaltered, covariant theory, then we see that there may be two types of differences:

(A) There may be effects of ordering of factors in products, such as the ones which may make $\mathcal{H}_\circ$ unequal $\tilde{\mathcal{H}}_\circ$ (compare the end of Sec. 12), or which may make $[F_\circ\ ;\ \mathcal{H}_\circ]/i\hbar c$ different from $[F_\circ, \mathcal{H}_\circ]$.

(B) There are in the expression (22.3), and therefore in (22.2), differences between $\tilde{\mathcal{H}}$ and the Hamiltonian $\tilde{\mathcal{H}}_\circ$ of the unaltered theory, causing a difference between the relation (22.2) between $F(t)$ and $F(0)$ in the altered theory, and the corresponding relation, covariant and unambiguous for physically meaningful quantities F, between their values at times t and 0 in the original theory.

The difference between $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}_\circ$ vanishes weakly. These weakly vanishing differences, though, occur as factors in various positions in the products of terms appearing in Eqs. (22.2)–(22.3). In the classical limit $\hbar\rightarrow 0$, however, (i.e., in a "quantum theory" in which effects of ordering of factors in most places would be overlooked), we could believe the assumption (4.6) according to which these products would vanish weakly irrespective of the position of the weakly vanishing factors. Under these circumstances, also the differences (A) should be overlooked. In this approximation, then, the time-dependence of the variables in our altered theory would be, at least "weakly," as covariant as in the unaltered theory.

We may also say that there is, in this approximation, covariance under the limitations imposed upon gauge and frames of reference by the auxiliary conditions. Under these conditions, indeed, the field equations are (weakly) covariant by the (weak) vanishing of the terms which are not generally covariant; and our classical approximation then ensures that integration of the conditionally (weakly) covariant field equations yields conditionally (weakly) covariant results.

In the "true" theory discussed in Sec. 24, we completely avoid the problems discussed here and avoid the necessity of a "classical approximation" by working there entirely with field variables which depend uniquely on time and with equations of motion which avoid dependence on weakly vanishing quantities.

Summarizing, in the altered theory the commutation relations are formally fully covariant, but the field equations are at best only weakly covariant, under the transformation group restricted by the weak conditions (16.5).

23. PARTICLE INTERPRETATION OF THE THEORY

In a chosen frame of reference as similar as possible to a rectangular Lorentz frame of reference and with the fourleg components as nearly equal to Kronecker symbols ($h_{(\alpha)}{}^\mu \approx \delta_\alpha{}^\mu$) as our curved space time admits, let $\tilde{q}_i(\mathbf{x})$ and $\tilde{p}^j(\mathbf{x})$ be the canonical field variables on a surface $x^0 =$ constant. We consider the field in a "cubical" region in which x^1, x^2, and x^3 each lie within an interval of size L, and in which we expand the field variables by

$$\tilde{q}_i(\mathbf{x}) \equiv V^{-\frac{1}{2}} \textstyle\sum_{\mathbf{k}} q_{i\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{x}) + c_i, \qquad (23.1a)$$

$$\tilde{p}^j(\mathbf{x}) \equiv V^{-\frac{1}{2}} \textstyle\sum_{\mathbf{k}} p^j{}_{\mathbf{k}} \exp(-i\mathbf{k}\cdot\mathbf{x}), \qquad (23.1b)$$

where $c_i \equiv \delta_\alpha{}^\mu$ for $q_i \equiv h_{(\alpha)}{}^\mu$, where $V = L^3$ and where $\mathbf{k}\cdot\mathbf{x}$ means $k_s x^s$. The summation over k_1, k_2, and k_3 is over positive and negative multiples of $2\pi/L$.

In our present case, there are "real" (i.e., Hermitian) Einstein-Bose fields satisfying

$$\tilde{q} \equiv \tilde{q}^*, \quad \tilde{p} \equiv \tilde{p}^*, \qquad (23.2)$$

and there are complex Fermi-Dirac fields for which

$$\tilde{p}^j \equiv i\hbar c \tilde{q}_j{}^*. \qquad (23.3)$$

For the real boson fields we put, with real constants f_j depending on $k = |\mathbf{k}|$ and to be determined later,

$$q_{j\mathbf{k}} \equiv q_{j(-\mathbf{k})}{}^* \equiv (\tfrac{1}{2}\hbar c)^{\frac{1}{2}} i f_j [l_{j\mathbf{k}} - l_{j(-\mathbf{k})}{}^*], \qquad (23.4a)$$

$$p^j{}_{\mathbf{k}}{}^* \equiv p^j{}_{(-\mathbf{k})} \equiv (\tfrac{1}{2}\hbar c)^{\frac{1}{2}} f_j{}^{-1} [l_{j\mathbf{k}} + l_{j(-\mathbf{k})}{}^*], \qquad (23.4b)$$

$$l_{j\mathbf{k}} \equiv (2\hbar c)^{-\frac{1}{2}} [f_j p^j{}_{\mathbf{k}}{}^* - i f_j{}^{-1} q_{j\mathbf{k}}], \qquad (23.4c)$$

so that

$$[l_{ik}\,;l_{jk'}^*] \equiv -[l_{jk'}^*\,;l_{ik}] \equiv \delta_{ij}\delta_{kk'}, \qquad (23.5a)$$

$$[l_{ik}\,;l_{jk'}] \equiv [l_{jk'}^*\,;l_{ik}^*] \equiv 0. \qquad (23.5b)$$

For the fermion fields, we have

$$[q_{ik}\,;q_{jk'}^*] \equiv (i\hbar c)^{-1}[q_{ik}\,;p^j{}_{k'}] \equiv \delta_{ij}\delta_{kk'}, \qquad (23.6a)$$

$$[q_{ik}\,;q_{jk'}] \equiv 0, \qquad (23.6b)$$

$$[q_{jk'}^*\,;q_{ik}^*] \equiv 0. \qquad (23.6c)$$

For the boson fields we define occupation numbers

$$N_{jk} \equiv l_{jk}^*\,l_{jk}. \qquad (23.7)$$

For the electron field, we perform a unitary transformation, from the Fourier components q_{ik} and q_{jk}^* of the undor components of the field, to linear combinations a_{rk} and a_{sk}^* of them, corresponding to flat-space free-electron states of positive and of negative energy. For the positive-energy states we define occupation numbers by $a_{rk}^*\,a_{rk}$, and for the negative-energy states we consider $a_{rk}a_{rk}^* \equiv b_{rk}^*\,b_{rk}$ an occupation number for a particle of opposite charge. (Here, $b_{rk} \equiv a_{rk}^*$.) From these definitions and (23.5)–(23.6) it follows that the occupation number operators have zero and integral eigenvalues only; only 0 or 1 for fermions. The l, the positive-energy a, and the b are annihilation operators, and their Hermitian conjugates are creation operators.

We now choose the $f_j(k)$. Writing out $\mathfrak{H}$, we find it to contain terms bilinear in the creation and annihilation operators, with coefficients which are functions of the c_i occurring in (23.1a). The sum of these terms we call the "unperturbed Hamiltonian," and the remaining terms in $\mathfrak{H}$ we call the perturbation. All the "nonlinearity" of the theory thus is part of the perturbation. By an appropriate choice of the $f_j(k)$ then we can give the unperturbed Hamiltonian the form of a sum of a meaningless vacuum energy term (to be subtracted) and a sum of $\pm(\hbar c|k_0|)\times$ the various occupation numbers, where $|k_0| = k$ for the graviton field and the photon field, and $|k_0| = +[\kappa^2+k^2]^{\frac{1}{2}}$ for the electron field. By introducing the $b_{rk} = a_{rk}^*$ we did already eliminate the minus signs from the unperturbed electron energy.

For the boson fields there remain minus signs. (For instance, the "scalar photon" energy term appears with a minus sign.[48]) One can show that these negative terms cancel against some positive terms on account of the auxiliary conditions,[48] or that the unperturbed energy is the sum of positive terms and weakly vanishing terms,[31] or one can introduce Gupta's indefinite metric and show that, on account of the auxiliary conditions, states which by themselves would have to be interpreted as negative-probability states occur in nature always combined with positive-probability states in such a way that the actual probability of the combined state vanishes in the unperturbed case.[49,52] In all these procedures, one may use the unperturbed ("linear") approximation of the auxiliary conditions, as we only want to show that the unperturbed particle energy is positive definite. In this approximation these problems have been solved in the past, except for the fourleg field with the fourleg conditions imposed, which would require a reformulation of Gupta's theory in terms of fourlegs.

The unperturbed theory is, of course, not generally covariant in the Heisenberg representation. The labeling of the particles by **k** has no covariant meaning. For discussion of the most interesting applications of the quantum theory of gravitation, this perturbation approach probably is completely useless. We refer here to such problems as the effect of the curvature of space on the self-energy of elementary particles. All we wanted to show in the present section was that, from the point of view of the field theoretician who considers gravitational effects negligible, the conventional treatment of the fields of photons and electrons remains possible, with a graviton field added to it as a next approximation; in as far as such a conventional treatment was justifiable in the first place.

PART III. UNSOLVED PROBLEMS

24. TRUE THEORY

One unsatisfactory feature of the altered theory was the unjustified use made over and over again of the assumption (4.6) that a product would be made weakly zero by a weakly vanishing factor whatever its position. It would be preferable to reformulate the theory in such a way that its formulas, as far as physically not meaningless, all would be strong equations. This, however, obviously requires an alteration of some commutation relations.

At the same time, this reformulation is to be used for eliminating from the fundamental formulas some quantities which show such ambiguities as dependence on the gauge of the Maxwell field. The following program of reformulation automatically seems to lead to variables which do not depend on the chosen frame of reference in space-time either, although, by relating the coordinates to the walls of and the apparatus in the laboratory, it would seem that an unambiguous meaning can be given to a chosen particular frame of reference. The fact that our present theory leads us more or less automatically to a description of the fields by invariants may be due to our failing to incorporate the laboratory and its gadgets into our Lagrangian, in an effort to keep our theory simple. If the laboratory would be included, distances between field points and laboratory apparatus could be described by invariants, making the theory applicable to the problems of the experimental physicist. How this would affect the formulation of the true theory requires investigation, as until now this problem has been neglected.

The key to the true theory lies in a canonical transformation defined by some of its properties, but until now not known in an explicit form. An example of

such a canonical transformation is the one proposed by Fermi for quantum electrodynamics in flat space,[45] in which the components of the vector potential in xyz space are transformed into the Fourier components of the longitudinal and the transverse components of this field, while the electric field-strength vector field is transformed into the Fourier components of the transverse electric field strength, and of the deviation of the longitudinal electric field strength from the Coulomb field. Such a canonical transformation is not local, although it is performed completely on a hypersurface $x^0 =$ constant. It may be convenient to execute it in several steps. If y stands for the variables $\tilde{q}$ and $\tilde{p}$, if $\mathbf{Z}$, $\mathbf{X}$, $\cdots$ are convenient parameters such as spatial coordinates or wave vectors or others, then such a transformation might for instance go through steps like

$$Y_a(\mathbf{Z},x^0) \equiv \int d^3\mathrm{x}\, F_a[\mathbf{Z}, \mathrm{x}, y_A(\mathrm{x},x^0)], \tag{24.1a}$$

$$\mathcal{Y}_\alpha(\mathbf{X},x^0) \equiv \int d^3\mathbf{Z}\, \mathfrak{F}_\alpha[\mathbf{X}, \mathbf{Z}, Y_a(\mathbf{Z},x^0)], \text{ (etc.).} \tag{24.1b}$$

Such a transformation is canonical from step to step if at each stage the $\mathcal{Y}_\alpha$ can be divided into two groups, the $\mathcal{P}^{\mathrm{i}}$ and the $\mathcal{Q}_{\mathrm{i}}$, in such a way that

$$[\mathcal{Q}_{\mathrm{i}}(\mathbf{X}), \mathcal{P}^{\mathrm{j}}(\mathbf{X}')] \equiv \delta_{\mathrm{i}}{}^{\mathrm{j}}\, \delta_3(\mathbf{X}-\mathbf{X}'), \tag{24.2a}$$

$$[\mathcal{Q}_{\mathrm{i}}(\mathbf{X}), \mathcal{Q}_{\mathrm{j}}(\mathbf{X}')] \equiv [\mathcal{P}^{\mathrm{i}}(\mathbf{X}), \mathcal{P}^{\mathrm{j}}(\mathbf{X}')] \equiv 0, \tag{24.2b–c}$$

where, for instance between steps one and two,

$$[\mathcal{A}(\mathbf{X}), \mathcal{B}(\mathbf{X}')] \equiv \int d^3\mathbf{Z} \left\{ \frac{\delta^P \mathcal{A}(\mathbf{X})}{\delta Q_{\mathrm{i}}(\mathbf{Z})} \frac{\delta^A \mathcal{B}(\mathbf{X}')}{\delta P^{\mathrm{i}}(\mathbf{Z})} - (-)^{u_A u_B} \frac{\delta^P \mathcal{B}(\mathbf{X}')}{\delta Q_{\mathrm{i}}(\mathbf{Z})} \frac{\delta^A \mathcal{A}(\mathbf{X})}{\delta P^{\mathrm{i}}(\mathbf{Z})} \right\}. \tag{24.3}$$

If in some step differential operators are involved, this can be expressed by using in (24.1) derivatives of delta functions. If we treat x, $\mathbf{Z}$, $\mathbf{X}$, $\cdots$ as indices, and the integrals over these variables as sums, the expressions (24.1) are simply polynomials or series in the field variables with all possible values of the indices, with coefficients which may depend on the values of the indices. In this sense we shall call the final variables $\mathcal{Y}$ or $\mathcal{Q}$ and $\mathcal{P}$ "polynomials" in the initial variables $\tilde{q}$ and $\tilde{p}$. The P.B. defined by (24.3) then are ordinary (modernized) P.B. with respect to the $\tilde{q}$ and the $\tilde{p}$, or, what is the same, modified P.B. with respect to the original q and p which followed from the Lagrangian before canonization. [Compare Eqs. (12.1) and following text.]

We now assume that it is possible to find a canonical transformation which leads to a set of $\mathcal{Q}$ and $\mathcal{P}$ such that some among the $\mathcal{P}$ will have the following two properties: The primary and secondary constraints [or, what amounts to the same, the *initial* auxiliary conditions (20.27)] shall be "depending throughout" on these particular $\mathcal{P}$ (which we shall distinguish from the other $\mathcal{P}$ by labeling them by an index f and by calling them the "fake" or the "phantom" momenta); and these $\mathcal{P}^{\mathrm{f}}$ shall be "depending throughout" on the constraints or initial auxiliary conditions. By saying that something "depends throughout" on a certain set of quantities we mean that it can be written as a "polynomial" (in our generalized sense) in such a way that *each* term in the polynomial contains *at least one* factor belonging to this certain set of quantities. Consequently, if this entire set of quantities would be set equal to zero, anything "depending throughout" on this set of quantities would also become equal to zero.

Our statement, therefore, really means that the two sets of equations

$$\mathcal{P}^{\mathrm{f}} = 0 \tag{24.4a}$$

and

$$\mathcal{C}^{\mathrm{n}} = 0 \tag{24.4b}$$

will follow from each other, if the $\mathcal{C}$ are the primary and secondary constraints or the initial auxiliary conditions.

There may be many sets of canonical variables $\mathcal{Q}$ and $\mathcal{P}$ of which some $\mathcal{P}$ have this property. We shall choose for the following discussions a set of variables for which the number of fake momenta $\mathcal{P}^{\mathrm{f}}$ is as small as possible. If we call the $\mathcal{P}$ which are no $\mathcal{P}^{\mathrm{f}}$ the "true momenta," and indicate them by an index t, then we make the number of true momenta $\mathcal{P}^{\mathrm{t}}$ as large as possible.

The question arises of what is the minimum number of fake momenta possible. In our theory there are sixteen constraints (24.4b). However, if one believes in the quantization of the altered theory, then these constraints are not really independent. In Sec. 11 we have seen that there exist relations between the modified P.B. of the constraints, and the constraints themselves. In particular, from Eq. (11.1a) one can solve for the $\Phi^{(r)(s)}$ in terms of modified P.B. of the $\Phi^{(r)(0)}$ and $\Phi^{(s)(0)}$. From (11.6d), one can solve for χ in terms of Φ and its modified P.B. with χ_0. From (11.6e) one can solve for the χ_r in terms of some other constraints and the modified P.B. of the Φ_r with χ_0. Now, if the quantization prescription (9.3) are believed, than we can express the modified P.B. in terms of commutators, which, after all, are differences of products and therefore are polynomials. Then, we could express the $\Phi^{(r)(s)}$, the χ_r, and χ, as quantities "depending throughout" on the $\Phi^{(r)(0)}$, the Φ_μ, Φ, and χ_0. *If*, therefore, we could express the latter nine constraints by functions depending throughout on say only nine $\mathcal{P}^{\mathrm{f}}$, then we would automatically get also the other seven constraints in terms of these nine $\mathcal{P}^{\mathrm{f}}$.

Notice that we cannot simply identify the $\mathcal{C}^{\mathrm{n}}$ themselves with the $\mathcal{P}^{\mathrm{f}}$, since the $\mathcal{C}^{\mathrm{n}}$ satisfy the relations of Sec. 11 for their modified P.B., while the $\mathcal{P}^{\mathrm{f}}$ satisfy the relations (24.2c). For the $\mathcal{P}^{\mathrm{f}}$ one should use some set of quantities depending throughout on the $\mathcal{C}^{\mathrm{n}}$, and at the same time satisfying (24.2c). We have not determined any set of such quantities as yet. However, it is probably impossible to express the nine independent constraints as polynomials depending throughout on less

than sixteen mutually commutative canonical (fake) field momenta $\mathcal{P}^f$. This would leave eight true momenta $\mathcal{P}^t$ for the fields of electrons, photons, and gravitons in interaction.

Let $\mathcal{Q}_f$ and $\mathcal{Q}_t$ be the canonical conjugates to the $\mathcal{P}^f$ and $\mathcal{P}^t$. The Hamiltonian $\tilde{\mathcal{H}}$, expressed as a "polynomial" in the new variables, will possess a part

$$\mathcal{H}_f \equiv \int d^3\mathbf{X}\, \mathcal{B}_f \mathcal{P}^f \quad (24.5)$$

depending throughout on the $\mathcal{P}^f$, and a part $\mathcal{H}_t$ completely independent of the $\mathcal{P}^f$:

$$\tilde{\mathcal{H}} \equiv \mathcal{H}_f + \mathcal{H}_t; \quad \delta\mathcal{H}_t/\delta\mathcal{P}^f(\mathbf{X}) \equiv 0. \quad (24.6a–b)$$

The coefficients $\mathcal{B}_f$ will in general be function(al)s of the $\mathcal{Q}_f$, $\mathcal{P}^f$, $\mathcal{Q}_t$, and $\mathcal{P}^t$. We shall now prove that $\mathcal{H}_t$ is not only independent of the $\mathcal{P}^f$, but also of the $\mathcal{Q}_f$.

The proof is based on the fact that the time derivatives of constraints depend throughout on the constraints. (See Secs. 11 and 20.) So, the equations (24.4b), valid at one time, must automatically be valid at all times. Then, the same must hold for the equivalent set of equations (24.4a), so that the $\mathcal{P}^f{}_{,0}$ are polynomials depending throughout on the set of $\mathcal{P}^f$.

On the other hand, the validity of Hamilton's equations in the altered theory tells us that

$$\mathcal{P}^f{}_{,0} \equiv -\, \delta^P\, \tilde{\mathcal{H}}/\delta\mathcal{Q}_f(\mathbf{X}). \quad (24.7)$$

Therefore, the $\delta^P\, \tilde{\mathcal{H}}/\partial\mathcal{Q}_f$ must depend throughout on the $\mathcal{P}^f$. This shows that in $\tilde{\mathcal{H}}$ the $\mathcal{Q}_f$ can occur only in terms containing some fake momentum, that is, in terms of the part $\mathcal{H}_f$ of the Hamiltonian. In other words,

$$\delta\mathcal{H}_t/\delta\mathcal{Q}_f(\mathbf{X}) \equiv 0. \quad (24.6c)$$

Up to there, we have merely discussed a canonical transformation in the unquantized altered theory. We now want to introduce the unquantized true theory.

The "true" theory is derived from the "pure" theory with the Langrangian $\tilde{L}^\circ(\tilde{q},\tilde{q}_{,\mu})$ of the end of Sec. 12, with field equations which in terms of the old variables q can also be derived from the basic Lagrangian $L_\circ(q,q_{,\mu})$. We define "pure" momenta by

$$\tilde{p}^{\circ i} \equiv \partial\tilde{L}^\circ/\partial\tilde{q}_{i,0}; \quad (24.8a)$$

$$p^{\circ i} \equiv \partial L_\circ/\partial q_{i,0}. \quad (24.8b)$$

Here, the $^\circ$ is the "purity index" indicating the use of the pure or basic Lagrangian. As compared to the momenta p of Parts I–II, which were knowingly (Sec. 18) or unknowingly (Secs. 3 and 14) derived from the altered Lagrangian, we have this time

$$\tilde{p}^{\circ i} \equiv \tilde{p}^i - \eta^i, \quad p^{\circ i} \equiv p^i - \eta^i. \quad (24.9a–b)$$

When we substitute the p° for the p in the constraints, we get (4.5). They become trivial identities $0 \equiv 0$ when one substitutes for the p° their definitions (24.8). Where we encounter P.B. of the constraints, though, we consider them functions of the q, $q_{,s}$, and p°:

$$\phi^{\circ k} \equiv \phi^k(p^\circ,q,q_{,s}); \quad (24.10a)$$

$$\phi^k(p^\circ(q,q_{,\mu}),\, q,\, q_{,s}) \equiv 0. \quad (24.10b)$$

In the pure theory, we do not distinguish strong and weak equations, so that in the following we shall simply write $=$ for $\equiv$, unless we want to indicate that the equality is a definition or a trivial identity.

We define "pure P.B." $(A,B)^\circ$ between functions A and B of the pure variables q and p°, by substituting differentiations with respect to p° for the $\partial/\partial p$ in (2.10). Replacing the P.B. in Eq. (9.4) by pure P.B., we define "pure modified P.B." $[A,B]^\circ$. The expressions of Sec. 10 for the modified P.B. remain valid for the pure modified P.B., if we replace all momenta p occurring in these formulas by p°. Again, pure modified P.B. are equal to pure regular P.B. with differentiations with respect to the canonized variables $\tilde{q}$ and $\tilde{p}^\circ$.

The pure Hamiltonian is given as a function of the $\tilde{p}^\circ$, $\tilde{q}$, and $\tilde{q}_{,\mu}$, by

$$\tilde{\mathcal{H}}^\circ{}_\circ \equiv \int\{\tilde{p}^{\circ i}\,\tilde{q}_{i,0} - \tilde{L}_\circ(\tilde{q},\tilde{q}_{,\mu})\}d^3\mathrm{x}. \quad (24.11a)$$

In the unquantized theory this is equal in value to

$$\mathcal{H}^\circ{}_\circ \equiv \int\{p^{\circ i}q_{i,0} - L_\circ(q,q_{,\mu})\}d^3\mathrm{x}, \quad (24.11b)$$

which is given by the expression for $\mathcal{H}_\circ$ in Sec. 6, with the p replaced by p°, and with the terms with θ constraints omitted if we use modified P.B. in the following. Also the term O_2 may be omitted, as by using p° we avoided the introduction of quantities η. [Contrary to the altered theory, the η are not introduced as chosen functions of the q and $q_{,\mu}$, so that their omission from the old formulas does *not* impose any auxiliary conditions $\eta(q,q_{,\mu}) = 0$.] Terms $\Phi^k B_k$ remain.

By the procedure of Sec. 11, we introduce the secondary constraints. The relations of Sec. 11 remain valid, if we add indices $^\circ$ behind the Hamiltonian and behind every momentum, every constraint, and every modified P.B.

The Hamiltonian equations (9.15) for any function F° of the q and p° may now be written as

$$F^\circ{}_{,0} = [F^\circ,\mathcal{H}^\circ{}_\circ]^\circ, \quad (24.12)$$

as the o by (7.3) will vanish with the η and ζ in the pure theory. This treating weak equations as strong equations makes, of course, quantization by (9.3) impossible; but we do not plan to quantize in such a simple way.

We now perform the canonical transformation outlined above, introducing "pure" fake and true variables $\mathcal{Q}^\circ$ and $\mathcal{P}^\circ$ which are functionals of the $\tilde{q}$ and $\tilde{p}^\circ$. From (24.12) and (24.2) we derive the Hamiltonian equations for the new variables. The Hamiltonian is the old one appearing in (24.5)–(24.7), except for purity indices on the variables and the omission of the mud terms O_2 and H_{add}, which in the altered theory did not contribute strongly to any P.B. or equations of motion

anyhow as they were quadratic in weakly vanishing expressions. In the pure theory, these terms do not appear in the first place.

As above, we prove that the true part of the pure Hamiltonian depends on the true pure variables only. The fake part of the Hamiltonian has the value zero on account of the vanishing of the fake momenta, but still might contribute to the P.B. of the Hamiltonian with the fake coordinates $\mathcal{Q}_{\mathrm{f}}^{\circ}$.

In the following, we omit the purity indices °, understanding them tacitly. We now prove that the *true* pure variables, $\mathcal{P}^{\mathrm{t}}$ and $\mathcal{Q}_{\mathrm{t}}$, are invariants under the transformations (20.4) further specified by (20.1)–(20.2) with (3.7), (3.9), and (3.12). Under such transformations of the original field variables q as functions of the coordinates $\mathbf{x},x^0$ (and not of the point P, as the latter changed to $\bar{P}$), the $q_{,\mu}$ are transformed by

$$\bar{\delta}(q_{,\mu}) = (\bar{\delta}q)_{,\mu}, \qquad (24.13)$$

and the transformation of the pure momenta is obtained by expressing them in terms of the q and the $q_{,\mu}$ by means of (24.8) [and *not* of (3.1) or (18.1)–(18.3)]. The transformed quantities $\mathcal{Q}$ and $\mathcal{P}$ are obtained by inserting in (24.1) for each step the transformed variables of the preceding stage in the right-hand member, but leaving the functional dependences, the values of the "coordinates" $\mathbf{x}$, $\mathbf{Z}$, $\mathbf{X}$, etc., and the intervals of integration unchanged.

In order to prove now the invariance of the true pure variables,

$$\bar{\delta}\mathcal{Q}_{\mathrm{t}} = 0, \qquad (24.14a)$$

$$\bar{\delta}\mathcal{P}^{\mathrm{t}} = 0, \qquad (24.14b)$$

we first construct the "generator" of the infinitesimal transformation group. It is given by

$$\mathcal{G} \equiv \int d^3\mathbf{x}\{\xi_{,0}\phi - \xi\chi + \tfrac{1}{2}\xi_{(\alpha)(\beta)}\Phi^{(\alpha)(\beta)} + \xi^{\mu}{}_{,0}\phi_{\mu} - \xi^{\mu}\chi_{\mu} - \xi^0(B\Phi + B_{(\alpha)(\beta)}\Phi^{(\alpha)(\beta)} + B^{\mu}\Phi_{\mu})\}, \qquad (24.15)$$

where again "purity" of the constraints is understood tacitly. The various quantities occurring in (14.15) are defined by (3.8), (3.13), (6.2), (6.3), (6.8), (6.10), (6.12), and (8.3), with indices °.

By straightforward calculation one then can verify the validity of the relations

$$\bar{\delta}F = [F,\mathcal{G}] \qquad (24.16)$$

for F equal to one of the various q and p of the theory, whence its validity for functionals of the q and p automatically follows. The formulas of Sec. 10 are used for evaluating the right-hand members of (24.16) for the various choices of F; in the resulting expressions, the pure momenta p are expressed by (24.8) in terms of the q and $q_{,\mu}$. The calculation is simplified by the vanishing of the constraints when not differentiated as in a P.B.; therefore, the modified P.B. of the products $B\Phi$ with F are simply $B[F,\Phi]$. Thus we are not bothered by the fact that the B depend on field velocities for which the P.B. are not known.

The calculation of the left-hand members of (24.16) for the various choices (q and p) for F was already discussed below (24.13).

If the two members of (24.16) for the various F would be calculated in the *altered* theory, using (18.1)–(18.3) instead of (3.1), one would find that the two members of (24.16) for some F would differ by amounts which would vanish only if the descriptors would satisfy the conditions (16.5). That is, in the altered theory only the *restricted* invariance transformations are generated by $\mathcal{G}$ according to (24.16). These restrictions were required for maintaining the validity of the auxiliary conditions and thus for maintaining the vanishing of the $\eta = \partial L_{\mathrm{add}}/\partial q_{,0}$, by which the momenta were altered. In the pure theory, there are no auxiliary conditions, and, correspondingly, there are no restrictions on the group of invariance transformations (20.4) for which (24.16) holds.

From the validity of (24.16) for the various q and p, its validity for any functionals of the q and p follows by (2.11)–(2.13). We may, therefore, apply (24.16) also to the $\mathcal{Q}$ and the $\mathcal{P}$. It is then convenient to express also the generator $\mathcal{G}$ in terms of the $\mathcal{Q}$ and the $\mathcal{P}$. From (24.15) we see that $\mathcal{G}$ depends throughout on the constraints $\mathcal{C}^n$; therefore, it also depends throughout on the (pure) fake momenta $\mathcal{P}^{\mathrm{f}}$. Since the latter vanish when not differentiated, the only contributions to $[F,\mathcal{G}]$ (and, therefore, to $\bar{\delta}F$) are the contributions from modified P.B. of F with the fake momenta $\mathcal{P}^{\mathrm{f}}$.

On the other hand, among all the $\mathcal{Q}$ and $\mathcal{P}$ the only ones having nonvanishing modified P.B. with the $\mathcal{P}^{\mathrm{f}}$ are the $\mathcal{Q}_{\mathrm{f}}$. Therefore, by (24.16), we find

$$\bar{\delta}\mathcal{P}^{\mathrm{f}} = 0, \qquad (24.17a)$$

$$\bar{\delta}\mathcal{P}^{\mathrm{t}} = 0, \qquad (24.17b)$$

$$\bar{\delta}\mathcal{Q}_{\mathrm{t}} = 0. \qquad (24.17c)$$

The result (24.17a) is trivial, as the canonical transformation leading to the $\mathcal{Q}$ and $\mathcal{P}$ had been chosen in such a way that the vanishing of the $\mathcal{P}^{\mathrm{f}}$ was simply a different way of stating the vanishing of the $\mathcal{C}^n$. Besides the trivial invariants $\mathcal{P}^{\mathrm{f}}$, however, we see from (24.17b) and (24.17c) that the true variables in the pure theory form a nontrivial set of general invariant functions of the "coordinates" $\mathbf{X}$ and x^0. Only the fake coordinates $\mathcal{Q}_{\mathrm{f}}$ are not invariant under the transformations (20.4).

As the $\mathcal{Q}_{\mathrm{f}}(\mathbf{X},x^0)$ depend on the arbitrary choice of coordinate system, gauge, and fourlegs, they cannot have a physical meaning *if* our theory is complete. (As stated above, it is *not*, since we forgot to include the laboratory system; and, therefore, an electron field in the south-west corner of the laboratory cannot be described in our theory by an invariant field function independent of the coordinate system, even though one would want to ascribe physical reality to such a field. We shall here ignore this fact, hoping that a more com-

plicated theory could be deviced for describing laboratory physics, and that a reasoning similar to the one given here could then be applied to that more complete theory.) Therefore, we regard the $\mathcal{Q}_f$ as *fake* variables indeed. The phantom momenta $\mathcal{P}^f$ are not even *variables*, as they are always zero. Only the $\mathcal{Q}_t$ and $\mathcal{P}^t$ then should be regarded as "true observables." [42]

It is, therefore, not only convenient but also justifiable to eliminate the fake variables completely from the theory. In our new canonical representation of the theory, this can appropriately be achieved by simply ignoring them. Doing so brings us from the "pure" theory into the true theory.

In the true theory, the pure $\mathcal{Q}_t$ and $\mathcal{P}^t$ are "from the beginning" the *only* variables. We now forget about the original Lagrangian with its q and the old Hamiltonian $\tilde{\mathcal{H}}^\circ{}_\circ$ with its $\tilde{q}$ and $\tilde{p}$, which were merely heuristic artifices. Our true theory is described by the knowledge of the Hamiltonian functional $\mathcal{H}_t[\mathcal{Q}_t(\mathrm{X}),\mathcal{P}^t(\mathrm{X})]$ introduced similar to (24.6a) in the pure theory, and independent of fake variables according to (24.6b) and (24.6c). The field equations of our true theory are, according to (24.12),

$$\mathcal{Q}_{t,0} = [\mathcal{Q}_t,\mathcal{H}_t] = \partial^A\mathcal{H}_t/\partial\mathcal{P}^t, \qquad (24.18a)$$

$$\mathcal{P}^t{}_{,0} = [\mathcal{P}^t,\mathcal{H}_t] = -\partial^P\mathcal{H}_t/\partial\mathcal{Q}_t. \qquad (24.18b)$$

The fact that for deriving the form of the functional $\mathcal{H}_t[\mathcal{Q}_t,\mathcal{P}^t]$ we needed (or at least used) the untrue theory with its $\tilde{q}$ and $\tilde{p}$ is a minor historical detail.

We are now ready to quantize the true theory. As there are no longer fake variables in the theory, we are no longer concerned about the constraints, which simply told us about the vanishing of the $\mathcal{P}^f$. As we ignore the $\mathcal{Q}_f$, we need not answer questions on whether they commute with the $\mathcal{P}^f$ or not (maybe they do). The only commutation relations which we need are the ones between the true variables. Just for them, we shall now maintain the validity of the quantization principle (9.3). Thus, we obtain the commutators

$$[\mathcal{Q}_t(\mathrm{X},x^0)\,;\mathcal{P}^{t'}(\mathrm{X}',x^0)] = i\hbar c\delta_t{}^{t'}\delta_3(\mathrm{X}-\mathrm{X}'), \qquad (24.19a)$$

$$[\mathcal{Q}_t\,;\mathcal{Q}_{t'}] = 0, \qquad (24.19b)$$

$$[\mathcal{P}^t\,;\mathcal{P}^{t'}] = 0. \qquad (24.19c)$$

The main difficulty with the true theory is that as yet the form of the canonical transformation leading from the $\tilde{q}(\mathbf{x})$ and $\tilde{p}(\mathbf{x})$ to the $\mathcal{Q}(\mathrm{X})$ and $\mathcal{P}(\mathrm{X})$ is not known, and that, consequently, such observables as the total energy-momentum four-vector of the system, known in terms of the former, are unknown in terms of the true variables. Some progress has been reported by DeWitt and by others about methods of finding this canonical transformation, partially using methods borrowed from the altered theory,[55,56] Some of the main troubles in this work are the nonvanishing of the modified P.B. between the constraints in their known form, and the dependence of the secondary coordinate constraints on products of canonical momenta.

The other main difficulty with the true theory is the necessity of incorporating in it some elements describing the setup of the initial state of an experiment in the laboratory, if we want to describe the experiment by some kind of true variables. This is equivalent to formulating a theory for a nonclosed system in which some external fields may be given.

25. INTERACTION REPRESENTATION

In the theory of fermions, it is necessary to distinguish particle states from antiparticle states. The electron wave-function $\tilde{\psi}$ may be interpreted as a superposition of negaton states multiplied by annihilation operators, and the complex conjugates of positon-state wave functions multiplied by creation operators. In the Heisenberg picture, an unambiguous and invariant distinction between these two types of states seems hard to formulate in a natural way.

The main merit of the discovery of the interaction picture in Lorentz-covariant flat-space field theory was that it enabled us to make this very distinction in an elegant way. It would, therefore, be useful to achieve the same in generally relativistic physics.

The major problem is to define the proper interaction operator W appearing in the generalized Schrödinger equation

$$i\hbar c\Psi_I(x^0)_{,0} = \int_{x^0=\text{constant}} j\, d^3\mathbf{x}\, W_I(\mathbf{x},x^0)\,\Psi_I(x^0). \qquad (25.1)$$

Here, the subscripts I denote the interaction picture. The operator $W(\mathbf{x},x^0)$ describes the local change of Ψ_I per unit four-volume near the point $(\mathbf{x},x^0)$, for instance as a bulge develops near that point in the space-like hypersurface $x^0 =$ constant. The interaction picture and the Heisenberg picture are related by a time-dependent unitary transformation $U(x^0)$, by

$$\Psi_I(x^0) = U(x^0)\Psi, \qquad (25.2a)$$

$$q_I(\mathbf{x},x^0) = U(x^0)q(\mathbf{x},x^0)U(x^0)^{-1}. \qquad (25.2b)$$

In the Heisenberg picture, $U(x^0)$ satisfies the equation

$$i\hbar cU(x^0)_{,0} = \int d^3\mathbf{x}U(x^0)W(\mathbf{x},x^0). \qquad (25.3)$$

A fact overlooked by many authors who wrote on this subject is that in general W is not a scalar, and that in general the commutator between W at two different points on the space-like hypersurface $x^0 =$ constant contains spatial derivatives of three-dimensional delta functions between those two points. Consequently, the interaction picture $q_I(\mathbf{x},x^0)$ of a field quantity $q(\mathbf{x},x^0)$ at a given world point will depend on the slope of the hypersurface $x^0 =$ constant through that point.[57] Thus, a quantity which is a scalar in the Heisenberg picture need not be a scalar in the interaction picture,

[56] F. J. Belinfante and D. I. Caplan, Bull. Am. Phys. Soc. Ser. II, 2, 12 (1957).

[57] F. J. Belinfante, Phys. Rev. 76, 66 (1949).

while the field $\tilde{\psi}$, which transforms in a horrible way in the Heisenberg picture, may well be expected to turn out to be an undor in the interaction picture of the altered theory, while the true variables describing the electron field may well turn out to be the components of a simple undor (4-component Dirac wave function) in the interaction picture of the true theory. At least, it is this way in the Lorentz-covariant true theory of the Maxwell field interacting with electrons, in so-called "gauge-independent quantum electrodynamics." [47]

Because of the dependence of $W_I(\mathbf{x},x^0)$ on the slope of the hypersurface $x^0=$constant through the point $(\mathbf{x},x^0)$, the integrability of (25.3) is not obvious. (It is trivial in Fermi's altered flat-space quantum electrodynamics, where W is a scalar, and where commutators $[W(\mathbf{x},x^0)\,;W(\mathbf{x}',x^0)]$ are simply proportional to the delta function $\delta_3(\mathbf{x}-\mathbf{x}')$; but it is by no means trivial in many other cases, like ours.) Therefore, whenever someone suggests a form for W, it is his task to prove that this choice for W makes (25.3) integrable; that is, given $U(x^0)$ on one hypersurface. the value of $U(x^0)$ on a later hypersurface $x^0=$constant should be proved to be independent of the choice of the coordinate system, the fourleg field, and the choice of gauge between these two hypersurfaces.[57,47]

In the Lorentz-covariant theory, it turns out that (25.3) in all cases investigated is integrable, if for W one chooses the interaction part of the T^{00} component of the symmetric energy-density tensor of the matter field. (In such theories, the matter energy is conserved by lack of gravitational energy.) On the other hand, many examples can be given of Lorentz-covariant theories in which the interaction part of the Hamiltonian density does *not* satisfy the integrability condition for (25.3).

For our gravitational theory, it is not known which definition of W will lead to an integrable equation (25.3), and at the same time will transform by (25.2)–(25.3) from the Heisenberg picture to an interaction picture in which the distinction between positon and negaton states is clear and in some sense invariant. The question arises whether the gravitational equations in such an interaction picture would reduce to the formulas of some flat-space theory of gravitation, like the linear approximation to Einstein's equations. The existence of such an interaction picture in the altered theory is a condition for the possibility of defining Gupta's indefinite metric for graviton states.[52] The interaction picture of the true theory would not require any indefinite metric.

A minor consequence of the existence of any interaction representation would be the possibility of formulating in a simple way a covariant perturbation procedure for the gravitational theory, which, as discussed at the end of Sec. 23, may be a rather useless tool.

Vol. 29
390A

REVIEWS OF MODERN PHYSICS

Volume 29, Number 4 — October, 1957

Synthesis of the Elements in Stars*

E. Margaret Burbidge, G. R. Burbidge, William A. Fowler, and F. Hoyle

Kellogg Radiation Laboratory, California Institute of Technology, and Mount Wilson and Palomar Observatories, Carnegie Institution of Washington, California Institute of Technology, Pasadena, California

"It is the stars, The stars above us, govern our conditions";
(*King Lear*, Act IV, Scene 3)

but perhaps

"The fault, dear Brutus, is not in our stars, But in ourselves,"
(*Julius Caesar*, Act I, Scene 2)

TABLE OF CONTENTS

* Supported in part by the joint program of the Office of Naval Research and the U. S. Atomic Energy Commission.

I. INTRODUCTION

A. Element Abundances and Nuclear Structure

MAN inhabits a universe composed of a great variety of elements and their isotopes. In Table I,[1] a count of the stable and radioactive elements and isotopes is listed. Ninety elements are found terrestrially and one more, technetium, is found in stars; only promethium has not been found in nature. Some 272 stable and 55 naturally radioactive isotopes occur on the earth. In addition, man has been able to produce artificially the neutron, technetium, promethium, and ten transuranic elements. The number of radioactive isotopes he has produced now numbers 871 and this number is gradually increasing.

Each isotopic form of an element contains a nucleus with its own characteristic nuclear properties which are different from those of all other nuclei. Thus the total of known nuclear species is almost 1200, with some 327 of this number known to occur in nature. In spite of this, the situation is not as complex as it might seem. Research in "classical" nuclear physics since 1932 has shown that all nuclei consist of two fundamental building blocks. These are the proton and the neutron which are called nucleons in this context. As long as energies below the meson production threshold are not exceeded, all "prompt" nuclear processes can be described as the shuffling and reshuffling of protons and neutrons into the variety of nucleonic packs called nuclei. Only in the slow beta-decay processes is there any interchange between protons and neutrons at low energies, and even there, as in the prompt reactions, the number of nucleons remains constant. Only at very high energies can nucleons be produced or annihilated. Prompt nuclear processes plus the slow beta reactions make it possible in principle to transmute any one type of

TABLE I,1. Table of elements and isotopes [compiled from *Chart of the Nuclides* (Knolls Atomic Power Laboratory, April, 1956)].

Elements		Isotopes	
Stable	81	Stable	272
Radioactive:		Radioactive:	
Natural ($Z\leq 83$)	1[a]	Natural ($A<206$)	11[d]
($Z>83$)	9[b]	($A\geq 206$)	44
Natural:		Natural:	
Stable and Radioactive	91	Stable and Radioactive	327
Radioactive:		Radioactive:	
Artificial ($Z\leq 83$)	1[c]	Artificial ($A<206$)	702
($Z>83$)	10	($A\geq 206$)	169
Total	102	Total	1198
Neutron	1	Neutron	1
	103		1199

[a] Tc, observed in *S*-type stars.
[b] Including At and Fr produced in weak side links of natural radioactivity
[c] Pm, not observed in nature.
[d] Including H^3, C^{14}, and Tc^{99}.

nuclear material into any other even at low energies of interaction.

With this relatively simple picture of the structure and interactions of the nuclei of the elements in mind, it is natural to attempt to explain their origin by a synthesis or buildup starting with one or the other or both of the fundamental building blocks. The following question can be asked: What has been the history of the matter, on which we can make observations, which produced the elements and isotopes of that matter in the abundance distribution which observation yields? This history is hidden in the abundance distribution of the elements. To attempt to understand the sequence of events leading to the formation of the elements it is necessary to study the so-called universal or cosmic abundance curve.

Whether or not this abundance curve is universal is not the point here under discussion. It is the distribution for the matter on which we have been able to make observations. We can ask for the history of that particular matter. We can also seek the history of the peculiar and abnormal abundances, observed in some stars. We can finally approach the problem of the universal or cosmic abundances. To avoid any implication that the abundance curve is universal, when such an implication is irrelevant, we commonly refer to the number distribution of the atomic species as a function of atomic weight simply as the atomic abundance distribution. In graphical form, we call it the atomic abundance curve.

The first attempt to construct such an abundance curve was made by Goldschmidt (Go37).† An improved curve was given by Brown (Br49) and more recently Suess and Urey (Su56) have used the latest available data to give the most comprehensive curve so far available. These curves are derived mainly from terrestrial, meteoritic, and solar data, and in some cases from other astronomical sources. Abundance determinations for

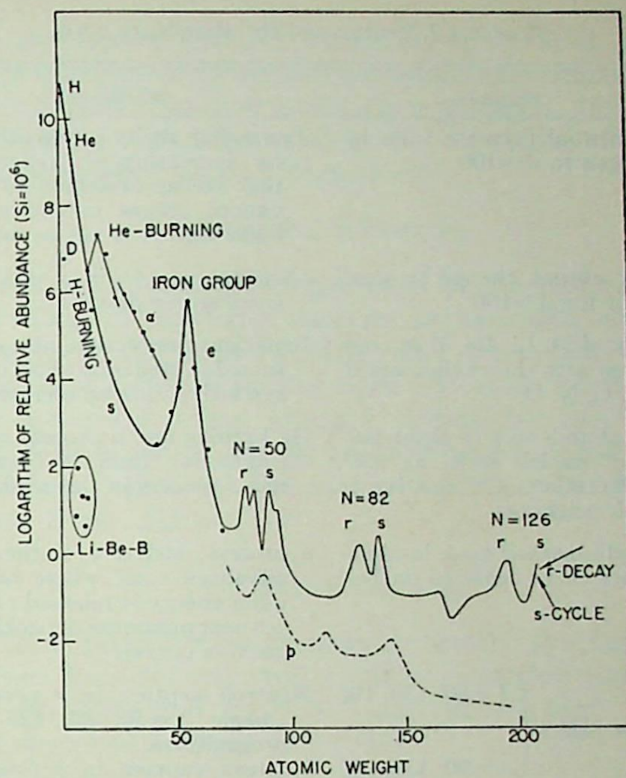


FIG. I,1. Schematic curve of atomic abundances as a function of atomic weight based on the data of Suess and Urey (Su56). Suess and Urey have employed relative isotopic abundances to determine the slope and general trend of the curve. There is still considerable spread of the individual abundances about the curve illustrated, but the general features shown are now fairly well established. These features are outlined in Table I,2. Note the overabundances relative to their neighbors of the alpha-particle nuclei $A=16, 20, \cdots 40$, the peak at the iron group nuclei, and the twin peaks at $A=80$ and 90, at 130 and 138, and at 194 and 208.

the sun were first derived by Russell (Ru29) and the most recent work is due to Goldberg, Aller, and Müller (Go57). Accurate relative isotopic abundances are available from mass spectroscopic data, and powerful use was made of these by Suess and Urey in compiling their abundance table. This table, together with some solar values given by Goldberg *et al.*, forms the basic data for this paper.

It seems probable that the elements all evolved from hydrogen, since the proton is stable while the neutron is not. Moreover, hydrogen is the most abundant element, and helium, which is the immediate product of hydrogen burning by the *pp* chain and the CN cycle, is the next most abundant element. The packing-fraction curve shows that the greatest stability is reached at iron and nickel. However, it seems probable that iron and nickel comprise less than 1% of the total mass of the galaxy. It is clear that although nuclei are tending to evolve to the configurations of greatest stability, they are still a long way from reaching this situation.

It has been generally stated that the atomic abundance curve has an exponential decline to $A\sim 100$ and is approximately constant thereafter. Although this is very roughly true it ignores many details which are important clues to our understanding of element synthesis. These details are shown schematically in Fig. I,1

† Refer to Bibliography at end of paper.

TABLE I,2. Features of the abundance curve.

Feature	Cause
Exponential decrease from hydrogen to $A \sim 100$	Increasing rarity of synthesis for increasing A, reflecting that stellar evolution to advanced stages necessary to build high A is not common.
Fairly abrupt change to small slope for $A > 100$	Constant $\sigma(n,\gamma)$ in s process. Cycling in r process.
Rarity of D, Li, Be, B as compared with their neighbors H, He, C, N, O	Inefficient production, also consumed in stellar interiors even at relatively low temperatures.
High abundances of alpha-particle nuclei such as O^{16}, $Ne^{20} \cdots Ca^{40}$, Ti^{48} relative to their neighbors	He burning and α process more productive than H burning and s process in this region.
Strongly-marked peak in abundance curve centered on Fe^{56}	e process; stellar evolution to advanced stage where maximum energy is released (Fe^{56} lies near minimum of packing-fraction curve).
Double peaks $A = 80, 130, 196$	Neutron capture in r process (magic $N = 50, 82, 126$ for progenitors).
Double peaks $A = 90, 138, 208$	Neutron capture in s process (magic $N = 50, 82, 126$ for stable nuclei).
Rarity of proton-rich heavy nuclei	Not produced in main line of r or s process; produced in rare p process.

and are outlined in the left-hand column of Table I,2. The explanation of the right-hand column will be given in Sec. II.

It is also necessary to provide an explanation of the origin of the naturally radioactive elements. Further, the existence of the shielded isobars presents a special problem.

B. Four Theories of the Origin of the Elements

Any completely satisfactory theory of element formation must explain in quantitative detail all of the features of the atomic abundance curve. Of the theories so far developed, three assume that the elements were built in a primordial state of the universe. These are the nonequilibrium theory of Gamow, Alpher, and Herman [see (Al50)], together with the recent modifications by Hayashi and Nishida (Ha56), the polyneutron theory of Mayer and Teller (Ma49), and the equilibrium theory developed by Klein, Beskow, and Treffenberg (Kl47). A detailed review of the history and development of these theories was given by Alpher and Herman (Al53).

Each of these theories possesses some attractive features, but none succeeds in meeting all of the requirements. It is our view that these are mainly satisfied by the fourth theory in which it is proposed that the stars are the seat of origin of the elements. In contrast with the other theories which demand matter in a particular primordial state for which we have no evidence, this latter theory is intimately related to the known fact that nuclear transformations are currently taking place inside stars. This is a strong argument, since the primordial theories depend on very special initial conditions for the universe. Another general argument in favor of the stellar theory is as follows.

It is required that the elements, however they were formed, are distributed on a cosmic scale. Stars do this by ejecting material, the most efficient mechanisms being probably the explosive ejection of material in supernovae, the less energetic but more frequent novae, and the less rapid and less violent ejection from stars in the giant stages of evolution and from planetary nebulae. Primordial theories certainly distribute material on a cosmic scale but a difficulty is that the distribution ought to have been spatially uniform and independent of time once the initial phases of the universe were past. This disagrees with observation. There are certainly differences in composition between stars of different ages, and also stars at particular evolutionary stages have abnormalities such as the presence of technetium in the S-type stars and Cf^{254} in supernovae. A detailed discussion of these and other features is given in Secs. XI and XII.

It is not known for certain at the present time whether all of the atomic species heavier than hydrogen have been produced in stars without the necessity of element synthesis in a primordial explosive stage of the universe. Without attempting to give a definite answer to this problem we intend in this paper to restrict ourselves to element synthesis in stars and to lay the groundwork for future experimental, observational, and theoretical work which may ultimately provide conclusive evidence for the origin of the elements in stars. However, from the standpoint of the nuclear physics alone it is clear that our conclusions will be equally valid for a primordial synthesis in which the initial and later evolving conditions of temperature and density are similar to those found in the interiors of stars.

C. General Features of Stellar Synthesis

Except at catastrophic phases a star possesses a self-governing mechanism in which the temperature is adjusted so that the outflow of energy through the star is balanced by nuclear energy generation. The temperature required to give this adjustment depends on the particular nuclear fuel available. Hydrogen requires a lower temperature than helium; helium requires a lower temperature than carbon, and so on, the increasing temperature sequence ending at iron since energy generation by fusion processes ends here. If hydrogen is present the temperature is adjusted to hydrogen as a fuel, and is comparatively low. But if hydrogen becomes exhausted as stellar evolution proceeds, the temperature rises until helium becomes effective as a fuel. When helium becomes exhausted the temperature rises still

further until the next nuclear fuel comes into operation, and so on. The automatic temperature rise is brought about in each case by the conversion of gravitational energy into thermal energy.

In this way, one set of reactions after another is brought into operation, the sequence always being accompanied by rising temperature. Since penetrations of Coulomb barriers occur more readily as the temperature rises it can be anticipated that the sequence will be one in which reactions take place between nuclei with greater and greater nuclear charges. As it becomes possible to penetrate larger and larger barriers the nuclei will evolve towards configurations of greater and greater stability, so that heavier and heavier nuclei will be synthesized until iron is reached. Thus there must be a progressive conversion of light nuclei into heavier ones as the temperature rises.

There are a number of complicating factors which are superposed on these general trends. These include the following.

The details of the rising temperature and the barrier effects of nuclear reactions at low temperatures must be considered.

The temperature is not everywhere the same inside a star, so that the nuclear evolution is most advanced in the central regions and least or not at all advanced near the surface. Thus the composition of the star cannot be expected to be uniform throughout. A stellar explosion does not accordingly lead to the ejection of material of one definite composition, but instead a whole range of compositions may be expected.

Mixing within a star, whereby the central material is mixed outward, or the outer material inward, produces special effects.

Material ejected from one star may subsequently become condensed in another star. This again produces special nuclear effects.

All of these complications show that the stellar theory cannot be simple, and this may be a point in favor of the theory, since the abundance curve which we are trying to explain is also not simple. Our view is that the elements have evolved, and are evolving, by a whole series of processes. These are marked in the schematic abundance curve, Fig. I,1, as H burning, He burning, α, e, r, s, and p processes. The nature of these processes is shown in detail in Fig. I,2; details of this diagram are explained in the following sections.

II. PHYSICAL PROCESSES INVOLVED IN STELLAR SYNTHESIS, THEIR PLACE OF OCCURRENCE, AND THE TIME-SCALES ASSOCIATED WITH THEM

A. Modes of Element Synthesis

As was previously described in an introductory paper on this subject by Hoyle, Fowler, Burbidge, and Burbidge (Ho56), it appears that in order to explain all of the features of the abundance curve, at least eight different types of synthesizing processes are demanded, if we believe that only hydrogen is primeval. In order to clarify the later discussion we give an outline of these processes here (see also Ho54, Fo56).

(*i*) *Hydrogen Burning*

Hydrogen burning is responsible for the majority of the energy production in the stars. By hydrogen burning in element synthesis we shall mean the cycles which synthesize helium from hydrogen and which synthesize the isotopes of carbon, nitrogen, oxygen, fluorine, neon, and sodium which are not produced by processes (ii) and (iii). A detailed discussion of hydrogen burning is given in Sec. III.

(*ii*) *Helium Burning*

These processes are responsible for the synthesis of carbon from helium, and by further α-particle addition for the production of O^{16}, Ne^{20}, and perhaps Mg^{24}. They are described in detail in Sec. III.

(*iii*) α *Process*

These processes include the reactions in which α particles are successively added to Ne^{20} to synthesize the four-structure nuclei Mg^{24}, Si^{28}, S^{32}, A^{36}, Ca^{40}, and probably Ca^{44} and Ti^{48}. This is also discussed in Sec. III. The source of the α particles is different in the α process than in helium burning.

(*iv*) *e Process*

This is the so-called equilibrium process previously discussed by Hoyle (Ho46, Ho54) in which under conditions of very high temperature and density the elements comprising the iron peak in the abundance curve (vanadium, chromium, manganese, iron, cobalt, and nickel) are synthesized. This is considered in detail in Sec. IV.

(*v*) *s Process*

This is the process of neutron capture with the emission of gamma radiation (n,γ) which takes place on a long time-scale, ranging from $\sim$100 years to $\sim 10^5$ years for each neutron capture. The neutron captures occur at a *slow* (s) rate compared to the intervening beta decays. This mode of synthesis is responsible for the production of the majority of the isotopes in the range $23 \leq A \leq 46$ (excluding those synthesized predominantly by the α process), and for a considerable proportion of the isotopes in the range $63 \leq A \leq 209$. Estimates of the time-scales in different regions of the neutron-capture chain in the s process will be considered later in this section, while the details of the nuclear physics of the process are discussed in Secs. V and VI together with the results. The s process produces the abundance peaks at $A=90$, 138, and 208.

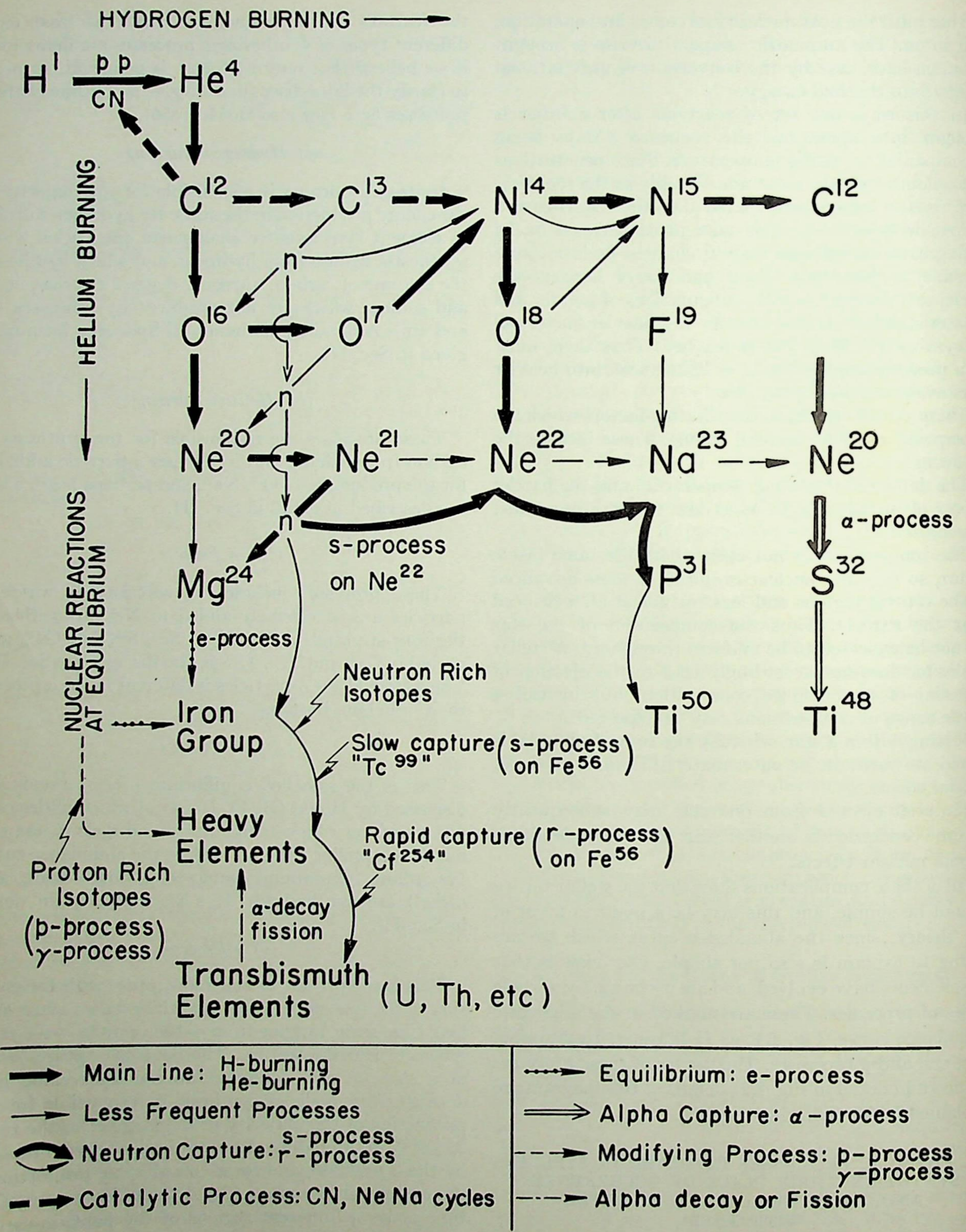


FIG. I,2. A schematic diagram of the nuclear processes by which the synthesis of the elements in stars takes place. Elements synthesized by interactions with protons (hydrogen burning) are listed horizontally. Elements synthesized by interactions with alpha particles (helium burning) and by still more complicated processes are listed vertically. The details of the production of all of the known stable isotopes of carbon, nitrogen, oxygen, fluorine, neon, and sodium are shown completely. Neutron capture processes by which the highly charged heavy elements are synthesized are indicated by curved arrows. The production of radioactive Tc^{99} is indicated as an example for which there is astrophysical evidence of neutron captures at a slow rate over long periods of time in red giant stars. Similarly Cf^{254}, produced in supernovae, is an example of neutron synthesis at a rapid rate. The iron group is produced by a variety of nuclear reactions at equilibrium in the last stable stage of a star's evolution.

(vi) r Process

This is the process of neutron capture on a very short time-scale, ~0.01–10 sec for the beta-decay processes interspersed between the neutron captures. The neutron captures occur at a *rapid* (*r*) rate compared to the beta decays. This mode of synthesis is responsible for production of a large number of isotopes in the range $70 \leq A \leq 209$, and also for synthesis of uranium and thorium. This process may also be responsible for some light element synthesis, e.g., S^{36}, Ca^{46}, Ca^{48}, and perhaps Ti^{47}, Ti^{49}, and Ti^{50}. Details of this process and the results of the calculations are discussed in Secs. VII and VIII. The *r* process produces the abundance peaks at $A=80$, 130, and 194.

(vii) p Process

This is the process of proton capture with the emission of gamma radiation (p,γ), or the emission of a neutron following gamma-ray absorption (γ,n), which is responsible for the synthesis of a number of proton-rich isotopes having low abundances as compared with the nearby normal and neutron-rich isotopes. It is discussed in Sec. IX.

(viii) x Process

This process is responsible for the synthesis of deuterium, lithium, beryllium, and boron. More than one type of process may be demanded here (described collectively as the *x* process), but the characteristic of all of these elements is that they are very unstable at the temperatures of stellar interiors, so that it appears probable that they have been produced in regions of low density and temperature. There is, however, some observational evidence against this which is discussed in Sec. X together with the details of the possible synthesizing processes.

In the upper half of Table II,1 the abundances of different natural groups of elements, taken from the atomic abundance table of Suess and Urey (Su56), have been summed and listed, firstly by number and by fraction of the total, and secondly by weight and by mass fraction of the total. In the lower half of Table II,1, a similar listing has been made, but in this case the isotopes have been grouped according to which mode of synthesis has been responsible for their production. Our method of assignment to these different processes is described in subsection B. In some cases a natural group comprises just the elements built by one process; e.g., the iron group of elements are also the *e*-process isotopes in the lower half of the table, and consequently the same entry appears in both halves. In one case, that of the *r*-process isotopes of intermediate atomic weight, the estimate given is one that has been calculated in Sec. VII.

An auxiliary but indispensable process which is also demanded in our description of element synthesis is a nuclear process which will provide a source of free neutrons for both the *s* process and the *r* process. The first suggestion in this direction was made by Cameron (Ca54, Ca55) and Greenstein (Gr54), who proposed that the $C^{13}(\alpha,n)O^{16}$ reaction would provide such a source. In addition to this the reaction $Ne^{21}(\alpha,n)Mg^{24}$ has also been proposed by Fowler, Burbidge, and Burbidge (Fo55). More detailed work on the rates of these reactions by Marion and Fowler is now available (Ma57) and some discussion is included in Sec. III.

B. Method of Assignment of Isotopes among Processes (i) to (viii)

Of the eight processes which are demanded to synthesize all of the stable isotopes, assignments among hydrogen burning, helium burning, the α process, the *e* process, and the *x* process are comparatively straightforward, and are implicit in previous work and in the discussions in Secs. III, IV, and X. Thus pure hydrogen

TABLE II,1. Atomic abundances of various groups of the elements from Suess and Urey (Su56).

Group	By number	Fraction of total	By weight	Fraction of total
H	4.00×10^{10}	0.928	4.03×10^{10}	0.755
He	3.08×10^{9}	0.071	1.23×10^{10}	0.231
Li, Be, B	1.44×10^{2}	3.3×10^{-9}	1.30×10^{3}	2.4×10^{-8}
Carbon group: C, N, O, Ne	4.01×10^{7}	9.3×10^{-4}	6.5×10^{8}	1.2×10^{-2}
Silicon group: Na−Sc	2.65×10^{6}	6.1×10^{-5}	7.3×10^{7}	1.3×10^{-3}
Iron group: $50\leq A\leq62$	6.4×10^{5}	1.5×10^{-5}	3.6×10^{7}	6.7×10^{-4}
Middleweight: $63\leq A<100$	1.1×10^{3}	2.6×10^{-8}	7.7×10^{4}	1.4×10^{-6}
Heavyweight: $A\geq100$	28	6.5×10^{-10}	4.6×10^{3}	8.6×10^{-8}
H+He burning: $12\leq A\leq22$	4.01×10^{7}	9.3×10^{-4}	6.5×10^{8}	1.2×10^{-2}
α process: 24, 28, ⋯48	2.2×10^{6}	5.1×10^{-5}	6.1×10^{7}	1.1×10^{-3}
s process: $23\leq A\leq46$	4.7×10^{5}	1.1×10^{-5}	1.3×10^{7}	2.4×10^{-4}
e process: $50\leq A\leq62$	6.4×10^{5}	1.5×10^{-5}	3.6×10^{7}	6.7×10^{-4}
s process: $63\leq A\leq75$	8.8×10^{2}	2.0×10^{-8}	5.7×10^{4}	1.1×10^{-6}
s process: $A>75$	1.1×10^{2}	2.6×10^{-9}	1.1×10^{4}	2.1×10^{-7}
r process: intermediate *A* (calculated)	$\sim1.5\times10^{3}$	$\sim3.5\times10^{-8}$	$\sim8\times10^{4}$	$\sim1.5\times10^{-6}$
r process: $A>75$	1.5×10^{2}	3.5×10^{-9}	1.4×10^{4}	2.6×10^{-7}
p process:	3.1	7.2×10^{-11}	2.0×10^{2}	3.8×10^{-9}

Summary: $X(\mathrm{H})=0.755$, $Y(\mathrm{He})=0.231$, $Z(A>4)=0.014$

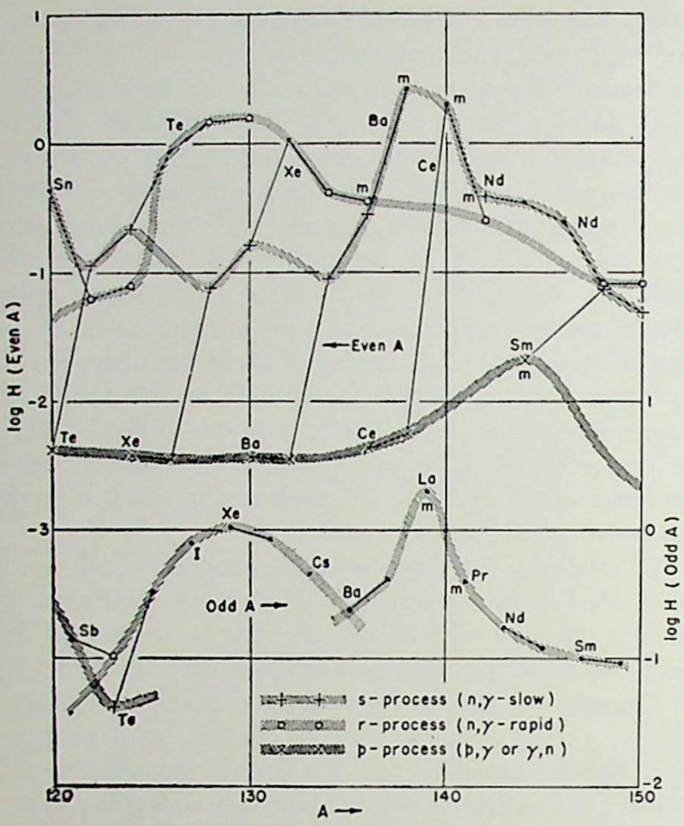


FIG. II,1. Logarithm of abundance H (silicon = 6.00, hydrogen = 10.60) of isotopes from $A = 120$ to 150 according to Suess and Urey (Su56). Odd mass numbers are shown in the bottom curve (read right ordinate). Even mass numbers are shown in the top curves (read left ordinate). Isotopes of a given element are connected by light lines (not to be confused with cross-hatched modes of element synthesis). Nuclei are distinguished by their method of synthesis as follows: +, produced only in neutron capture at a slow rate (*s* process); ○, produced only in neutron capture at a rapid rate (*r* process); ●, produced in both processes but predominant mode of production assigned as discussed in text; ×, produced only in proton capture or photoneutron processes (*p* process). The paths of the three processes are then indicated by the cross-hatching indicated in the figure. The stable nuclei with the magic number of closed shell neutrons, $N = 82$, are indicated by *m*. The abundance peak near $A = 139$ follows from the low neutron-capture cross section of the magic stable nuclei in the *s* process. The abundance peak near $A = 129$ follows from the low neutron-capture cross section and slow beta decay in the *r* process for the magic neutron-rich isobars (for example, $A = 129$, $N = 82$, $Z = 47$) which eventually decay by beta emission to the stable nuclei in this region. These arguments are based on the fact that in "steady flow" the abundance of a given nucleus will be inversely proportional to the rate at which it is transmuted by neutron capture or beta decay. The alternation in abundances exhibited by the sequence Te^{122}, Te^{124}, Xe^{128}, Xe^{130}, and Ba^{134}, Ba^{136} is to be expected in the *s* process because of the difficulty in adding further neutrons after two have already been captured. In general there is only one isobar at odd mass numbers. However, Te^{123} and Sb^{123} differentiate the *r* and *s* processes in an odd *A* curve. Te^{123} is probably unstable but with a very long lifetime.

burning is responsible for the synthesis of helium, and pure helium burning is responsible for C^{12}, O^{16}, and Ne^{20}. Also, hydrogen burning in these latter products of helium burning accounts for all of the stable isotopes up to Ne^{22} (cf. Sec. III). The only exceptions up to this point are deuterium, lithium, beryllium, and boron, which have therefore been assigned to a radically different type of synthesizing process, the *x* process. Assignment to the α process for a number of four-structure nuclei was originally suggested by Hoyle (Ho54). All of these nuclei lie in a region of the abundance curve in which the surrounding nuclei are synthesized by the *s* process. As described in detail in Sec. VI, a plot of the σN products (neutron capture cross section times abundance) for such nuclei should delineate a smooth curve. However, because of their comparatively large abundances, the nuclei made by the α process all lie above this curve. This result bears out the assignment of these nuclei to the α process. Assignment of the nuclei in the iron peak to the *e* process was also proposed earlier by Hoyle. The large abundance peak at this point shows clearly that a separate process is demanded and the results of Sec. IV suggest strongly that our assignments to the *e* process are correct.

As far as assignments among the *s*, *r*, and *p* processes are concerned, the situation is a little more complex. Suess and Urey (Su56) and Coryell (Co56) have already pointed out that the peaks in the abundance curves at stable nuclei with filled neutron shells ($A = 90$, $N = 50$; $A = 138$, $N = 82$; $A = 208$, $N = 126$) strongly indicate the operation of the *s* process, and the nearby peaks at $A = 80$, 130, and 194, shifted by $\delta A \sim 8$ to 14, similarly require the operation of the *r* process. Also, calculations of Fowler *et al.* (Fo55) suggested that apart from the nuclei built by the α process all of the nuclei with $23 \leq A \leq 46$ with the exception of S^{36}, Ca^{46}, and Ca^{48} can be synthesized by the *s* process.

Certain isotopes of the heavy elements can be built only by the *r* process while others can be built only by the *s* process. The two processes differ in this respect because they allow very different times for the occurrence of the beta disintegrations along the chain of nuclei built by the neutron addition. In the *s* process most beta-active nuclei have time to decay before additional capture occurs. On the other hand, the *r* process involves neutron captures which take place at a rate much faster than the beta decays which have reaction times of 0.01 to 10 sec.

The need for a third process, the *p* process, arises because a third set of isotopes cannot be built by either the rapid or the slow capture of neutrons (the reasons for this from the standpoint of nuclear structure are discussed in Sec. V). As an example of the way in which the three processes operate and also of how assignments have been made, we shall consider the eight isotopes of tellurium. The light isotopes ${}_{52}Te_{70}^{122}$, ${}_{52}Te_{71}^{123}$, and ${}_{52}Te_{72}^{124}$ can be synthesized only in the *s* process, since in the *r* process, the ultimate beta decays of the neutron-rich isobars at 122, 123, 124, produced by rapid neutron addition, terminate at the stable isotopes ${}_{50}Sn_{72}^{122}$, ${}_{51}Sb_{72}^{123}$, and ${}_{50}Sn_{74}^{124}$, which are on the neutron side of the nuclear stability line. On the other hand, the heaviest isotope Te^{130} can be produced only in the *r* process, where it is the stable product of the decay of neutron-rich isobars of mass 130. In the *s* process, radioactive Te^{129} with a half-life of 70 min has time to decay to I^{129} (half-life 2×10^7 years), and after another neutron

capture, the resultant I^{130} decays in 12.6 hr to Xe^{130} which is thus produced in the *s* chain instead of Te^{130}. The isotopes Te^{125}, Te^{126}, and Te^{128} can be produced in either the *s* or the *r* process, although Te^{128} is produced in the slow capture of neutrons only in a weak side link of the chain resulting from the fact that I^{128} decays 5% of the time by positron emission or electron capture. Thus we believe that it is synthesized predominantly by the *r* process. The rarest and lightest isotope Te^{120} cannot be built by either process and it is for this isotope that the *p* process is demanded. Te^{120} is about 1% as abundant, and Te^{122}, Te^{123}, and Te^{124} are about 10% as abundant, as Te^{128} and Te^{130}. This suggests that we assign Te^{126}, which has an abundance comparable to Te^{128} and Te^{130}, to the *r* process. Te^{125} is an intermediate case, but it follows the trend of the *r* process and to this we assign its production. Assignments between the *s*, *r*, and *p* processes have been made in this way. In Figs. II,1; II,2; and II,3, the separation of the isotopes in the region $120<A<150$ is shown. In Fig. II,1 the abundances are plotted logarithmically after the manner of Suess and Urey (Su56). Nuclear species produced in the same process are connected by shaded curves and the general trend of the production becomes clear. In order to show more clearly the great increase in abundance in the peaks, linear plots of the abundances of the odd and even isotopes, respectively, are shown in Figs. II,2 and II,3. The magic-number peaks stand out clearly in both odd-A and even-A nuclei.

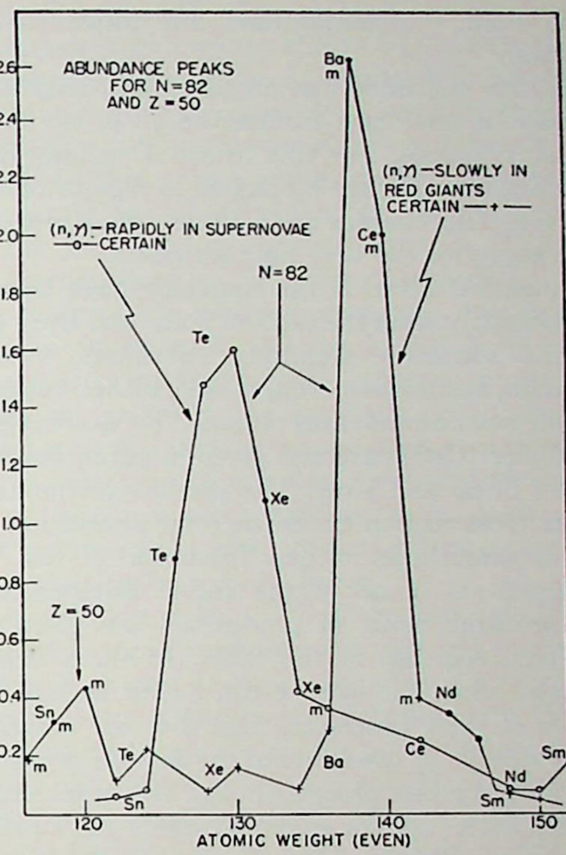


FIG. II,3. The even-A abundance peaks near $A=130$ and 138 shown on a linear scale. See Fig. II,1 for comments.

C. Abundances and Synthesis Assignments Given in the Appendix

The Appendix contains all of the information we have been able to collect which is relevant to the synthesis

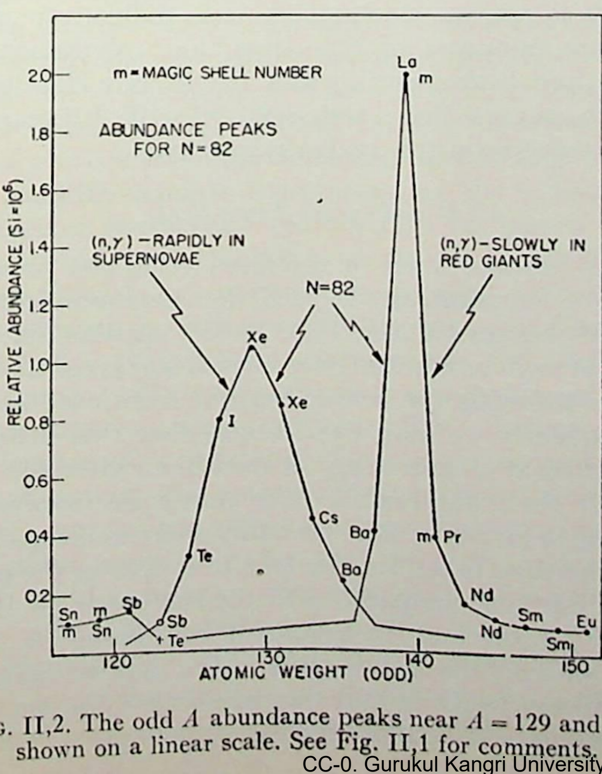


FIG. II,2. The odd A abundance peaks near $A=129$ and 139 shown on a linear scale. See Fig. II,1 for comments.

problem. All of the stable isotopes in order of increasing A are given in this table. We also include the beta-unstable isotopes which lie on the main neutron-capture chain in the *s* process, together with a few others which are given for special reasons. The table was originally drawn up to represent the building of elements by the *s* process. For this reason the left-hand columns follow the main chain of nuclei synthesized by neutron capture in the *s* process. The right-hand columns give information concerning the isotopes which are either completely by-passed by the *s* process or which lie in the subsidiary loops of the chain. These loops can be formed in two ways. Either they form at nuclei which beta decay with a half-life such that in the *s* process a proportion of the nuclei will capture a further neutron before beta decay, while the remainder beta decay directly (the next step in the chain following the nucleus $X_Z{}^A$ being either $X_Z{}^{A+1}$ or $X_{Z+1}{}^A$). Alternatively a loop will be formed if a nucleus can decay either by emitting an electron or by emitting a positron or capturing a K electron (the next step in the chain following the nucleus $X_Z{}^A$ being either $X_{Z+1}{}^A$ or $X_{Z-1}{}^A$). A few isomeric states are also given [an example is ${}_{48}Cd_{65}{}^{(m)}(\beta^-)$] in cases in which they also lead to the development of loops. In all cases the side of the loop which is in the main chain will be determined by the faster of the reactions (either neutron-capture or beta-decay) which can take place. To distinguish between nuclei which lie in the weak branch of a loop, and those which are completely by-passed

in the chain, horizontal lines are drawn to contain each loop.

Although the table was prepared primarily for the *s* process it is now used to describe all of the modes of element synthesis. For this reason a number of beta-active nuclei which are by-passed in regions of small *A* are given. This enables us to discuss the *r* process as it would apply for the very light elements.

Abundances given in the Appendix have been taken predominantly from the table of Suess and Urey (Su56). For the elements scandium, titanium, vanadium, chromium, manganese, iron, cobalt, nickel, copper, and lead, the solar abundances obtained by Goldberg, Aller, and Müller (Go57) are also given in parentheses under those of Suess and Urey. The product of the neutron-capture cross section times the solar abundance is also given in parentheses in the σN column (cf. Sec. VI).

Assignments of all of the stable isotopes to their most probable mode of production are given in the right-hand columns in the table. In the *s* and the *r* processes, when the isotope can be made only by one or other of these processes, then they are designated *s only* or *r only.* In cases where the isotope is made predominantly by one process though the other may contribute a little—as, for example, where the *r* process is mainly responsible for synthesis, though the isotope lies in the main *s* chain—then we write simply *r* (or *s*), *hydrogen burning*, *helium burning*, α, *e*, *p*, *x*, as the case may be. In some cases more than one process may contribute appreciably to the synthesis of that isotope. In these cases the process which is believed to contribute most is written first, while in the few cases in which the two processes are believed to make equal contributions, an approximate equality sign is used, and in the abundance column ($\frac{1}{2}$) is written. In some other cases $\frac{2}{3}$ of the abundance of an isotope has been attributed to the *s* process and only $\frac{1}{3}$ to the *r* process. In these cases ($\frac{2}{3}$) is written in the abundance column (these fractions are only very approximate). Thus, in proceeding from isotopes which can be made solely on the *s* process to those which are made solely on the *r* process, the sequence of designations is *s only*, *s*, *sr*, $s\approx r$, *rs*, *r*, *r only.* Magic-number nuclei with closed shells having 14, 20, 28, 50, 82, and 126 neutrons, most of which are synthesized by the *s* process, are marked in the assignment column of the table by *m* in parentheses. Similarly, nuclei made by the *r* process whose progenitors had closed shells having 50, 82, or 126 neutrons are marked by *m′* in parentheses. A few of the light nuclei undergo nuclear reactions which lead not to heavier element synthesis but to breakup, or to decrease in *A* (e.g., Li^9). In the table these have been designated *return* (meaning that they return back down the chain), and in some cases the products of the reactions which are responsible for this are given.

D. Time-Scales for Different Modes of Synthesis

(1) *Hydrogen-Burning Time-Scale*

Hydrogen burning is responsible for all of the energy production of stars on the main sequence. After stars have evolved off the main sequence following the development of a chemical inhomogeneity, hydrogen burning in a shell still remains an important energy source. Thus the synthesis of elements in hydrogen burning is going on continuously, and the range in time-scales for particular stars is dependent only on their initial masses after condensation, and the point in their evolution at which they eject material which has been synthesized. Thus these time-scales may range from $\sim 10^6$ years for massive *O* and *B* stars to times which are of the order of though less than the age of the galaxy (since we have no evidence for the existence today of primeval stars of pure hydrogen).

(2) *Helium-Burning Time-Scale*

It is believed that helium burning takes place in stars which have evolved onto the giant branch of the *HR* diagram. In this region central temperatures of $\sim 10^8$ degrees and densities of $\sim 10^5$ g/cc are reached in the helium core, according to the theoretical calculations of Hoyle and Schwarzschild (Ho55), and under these conditions synthesis of C^{12} becomes possible (Sa52, Co57, Sa57). The further helium-burning reactions leading to production of O^{16}, Ne^{20}, and Mg^{24} also can take place under suitable conditions though the amounts of these isotopes synthesized will decrease as the Coulomb barriers get larger (cf. Sec. III). The time-scale for helium burning is therefore governed by the lifetime of stars subsequent to their becoming red giants. Since their mode of evolution off the giant branch is not yet understood, it is difficult to make accurate estimates of this time, but calculations of Hoyle and Schwarzschild and Hoyle and Haselgrove (Ho56a) suggest that a time-scale of $\sim 10^7$–10^8 years for stars to evolve in this region is reasonable.

(3) *α-Process Time-Scale*

This process, since it involves reactions between nuclei of comparatively large *Z*, demands temperatures of $\sim 10^9$ degrees (cf. Sec. III). Thus it appears that this condition will be reached only when a star is contracting further following the helium burning mentioned above. Consequently we may hypothesize that this situation is reached at a later stage of evolution, after the star has left the giant branch. The time-scale involved is difficult to estimate. It is probably short in comparison with the time taken for the star to evolve in the giant region but long compared with the time-scale for the *e*, *r*, and *p* processes. An argument concerning the synthesis of Ti^{48}, the last nucleus which may be built by this process, suggests that the time-scale for its production from the preceding nucleus is $\gtrsim 20$ years. Details

of this argument are given in Sec. III D. The total time-scale for the whole α process might lie in the range $\sim 10^2$–10^4 years.

(4) *Rapid Time-Scales*

e process.—As is discussed in Secs. IV and XII, this process takes place under extreme conditions of temperature and density, probably just prior to a supernova outburst. The time-scale involved may be of the order of seconds or minutes.

r process.—The nuclear physics of this process demands that neutrons be added extremely rapidly, so that the total time-scale for the addition of a maximum of about 200 neutrons per iron nucleus is $\sim$10–100 sec. It has previously been suggested that the spontaneous fission of Cf^{254} is responsible for the form of the decay light curves of supernovae of Type I which have an exponential form with a half-life near 55 days, so that this is strong evidence for the fact that the *r* process takes place in such outbursts (Bu56). The time-scale for the explosive phases of such outbursts may well be as short as 100–1000 sec. Further support is given by a comparison of the results of calculations based on this hypothesis with the observed abundances of isotopes built by this process (Secs. VII and VIII).

p process.—It has previously been suggested that this process takes place in the outbursts of supernovae of Type II on a time-scale comparable to that for the *r* process.

(5) *s-Process Time-Scale*

In earlier work it was suggested that the *s* process is currently going on in the interiors of some red giant stars (Ca55, Fo55). The presence of technetium in the atmospheres of *S*-type stars, observed by Merrill (Me52), and also to a lesser extent in carbon and *M*-type stars, affords conclusive evidence that nuclear activity involving neutron capture is currently going on in the interiors of these stars. This element is being both synthesized and mixed to the surface of the star in a time which is less than or at least of the same order as the half-life against beta decay of the longest-lived isotope of technetium (Tc^{99} with a half-life of 2×10^5 years). This is compatible with the result that the time taken for the stars to evolve in the giant branch of the *HR* diagram is $\sim 10^7$ years. Of the reactions $C^{13}(\alpha,n)O^{16}$ and $Ne^{21}(\alpha,n)Mg^{24}$, which can give neutron production, the former will give an adequate supply of neutrons to carry heavy-element synthesis right through to the heaviest nuclei only if there is some mixing between the envelope hydrogen and the products of helium burning in the cores at the red-giant stage. In a previous paper (Fo55) it was pointed out that the $Ne^{21}(\alpha,n)Mg^{24}$ reaction could provide an adequate source of neutrons and also avoid the problem of mixing between core and envelope. In Sec. III F, however, it is pointed out that for this to be the case, certain conditions must be satisfied. The relative importance of these two reactions for the *s* process will be discussed in detail in Sec. III F. The evidence from work on theoretical models suggests that a very deep convective zone extending inward from the surface is not unreasonable, so that mixing of the synthesized material to the surface can take place.

As will be seen, there appear to be two time-scales for the *s* process. These depend on the average times taken for the nuclei to capture neutrons. Thus these times are functions of the neutron density and hence of the rates of the neutron-producing reactions, and also of the capture cross sections. In Sec. V we describe the method of estimating the neutron capture cross sections (σ) given in the appendix. By using these together with the isotopic abundances and the assignments of nuclei to the *s* process, it is possible to make some estimates of the neutron-capture times at particular points in the slow neutron-capture chain. The method of doing this is as follows.

As pointed out in Sec. II C, if the beta-decay time is greater than a few years then the next link in the chain following X_Z^A may be either X_{Z+1}^A or X_Z^{A+1}, the relative importance of the branch depending on the time-scale for neutron capture at this point. We show in the appendix all possible loops corresponding to time-scales between $\sim 10^2$ years and $\sim 10^6$ years. Thus in all cases in which the beta decay takes place in a time $\ll 100$ years we have assumed that all of the nuclei will beta decay and any nuclei which can only be produced if another neutron is captured before the beta decay takes place are by-passed in the *s* process. If on the other hand the half-life for beta decay is $\gg 10^6$ years we have assumed that the nucleus is effectively stable in the *s* chain and will always capture a neutron. In cases like this it must be remembered, however, that all of the unstable nuclei with half-lives $\gg 10^6$ years which are effectively stable as far as the *s* process is concerned will eventually decay into the succeeding stable nuclei long after the *s* process has ceased operating. For elements in the solar system this will be true for all long-lived but unstable nuclei for which the half-life $<5\times10^9$ years.

There appear to be two indicators for the time-scale for the *s* process in two regions of *A*. The first can be obtained from two isotopes of krypton. Both Kr^{80} and Kr^{82} are built only by the *s* process, being shielded in the *r* process by Se^{80} and Se^{82}. However, from the appendix we see that there is an *s* chain loop at Se^{79} which is due to the beta decay of this isotope with a half-life of 7×10^4 years. Thus while Kr^{82} is produced partly in the main chain and partly in the loop, Kr^{80} is only produced in the loop. Now under conditions of steady neutron flow (see Sec. VI) the basic equations for the relative abundances of these isotopes are

$$N(Kr^{80})\lambda_n(Kr^{80})=0.92N(Se^{79})\lambda_\beta(Se^{79})$$

and

$$N(Kr^{82})\lambda_n(Kr^{82})=N(Se^{79})[\lambda_\beta(Se^{79})+\lambda_n(Se^{79})],$$

where λ_n and λ_β are the neutron-capture and beta-decay rates respectively. The coefficient 0.92 appears because in the loop the nuclei have to pass through Br^{80} which beta decays only 92% of the time. On dividing these two equations and substituting in terms of the half-lives t_n and t_β instead of the rates, we find that

$$\frac{t_n(\mathrm{Se}^{79})+t_\beta(\mathrm{Se}^{79})}{t_n(\mathrm{Se}^{79})}=0.92\frac{t_n(\mathrm{Kr}^{80})}{t_n(\mathrm{Kr}^{82})}\frac{N(\mathrm{Kr}^{82})}{N(\mathrm{Kr}^{80})}.$$

Now $\sigma(\mathrm{Kr}^{80})\simeq2\sigma(\mathrm{Kr}^{82})$ so that $\lambda_n(\mathrm{Kr}^{80})\simeq2\lambda_n(\mathrm{Kr}^{82})$ and $t_n(\mathrm{Kr}^{82})\simeq2t_n(\mathrm{Kr}^{80})$. [In the appendix we put $\sigma(\mathrm{Kr}^{80})=\sigma(\mathrm{Kr}^{82})$ and do not in general take into account the factor of ~2 which is a result of the fact that the light neutron-poorer isotopes capture neutrons more readily than the heavier ones.] The abundance ratio $N(\mathrm{Kr}^{82})/N(\mathrm{Kr}^{80})$ is given in the appendix. Thus substituting these values on the right-hand side of the equation we find for Se^{79} that

$$t_n/t_\beta=1/1.37,$$

and, since $t_\beta=7\times10^4$ years, $t_n=5.1\times10^4$ years. Now this half-life for neutron capture is determined in a region where $\sigma_n=360$ mb, so that in the region near iron where this neutron chain begins and where $\sigma=30$ mb the effective half-life is $5.1\times10^4\times360/30\simeq6\times10^5$ years. The total time demanded to build up the isotopes from iron if this were characteristic of the whole curve would be 10^6–10^7 years since there are many captures involved, but most of these take place on nuclei with larger σ values than that for iron. We show later that this long time-scale for the *s* process is probably the result of a paucity of neutrons—probably only about five neutrons per iron nucleus are made available. This is reflected in the tendency for the σN product to decrease rather rapidly as A increases.

The second indicator for an *s* process time-scale comes from consideration of the isotopes of gadolinium at $A=152$ and 154. Of these isotopes Gd^{154} can only be made by the *s* process. On the other hand the majority of Gd^{152} is probably made in the *p* process (Sec. IX), since its abundance fits in well with the run of abundances of the isotopes made by this process in this region of the curve. We shall arbitrarily suppose that of the total abundance of Gd^{152} of 0.00137 about 0.001 is made by the *p* process (cf. Fig. IX,1) and the remainder by the *s* process, since the following argument is in no way critically dependent on this division. The small amount made in the *s* process must be due to the fact that Gd^{152} is mostly by-passed in the *s*-process chain. Reference to the appendix shows that the branching which does this cannot take place at Eu^{152} since in 28% of the cases Eu^{152} will decay to Gd^{152} while in 78% of the cases isomeric Eu^{152m} will decay to Gd^{152}. The combination of these two decays shows that in 36% of the cases Eu^{152} will decay to Gd^{152} if the production ratio $Eu^{152}/Eu^{152m}\sim5$. Thus the branching responsible for this by-pass must take place at Sm^{151} which has a half-life for beta decay of 80 years and which must mainly capture a neutron before beta decay. Thus we must have that

$$\frac{t_n(\mathrm{Sm}^{151})}{t_\beta(\mathrm{Sm}^{151})}=\frac{N'(\mathrm{Gd}^{152})}{N(\mathrm{Gd}^{154})},$$

where $N'(\mathrm{Gd}^{152})$ is the abundance of Gd^{152} estimated on the supposition that 36% of the decays from Eu^{152} give an abundance of 0.00037. That is, $N'(\mathrm{Gd}^{152})=0.00037/0.36\simeq0.001$, and $N(\mathrm{Gd}^{154})$ is the abundance of this isotope given in the appendix. Thus substituting for $N'(\mathrm{Gd}^{152})$, $N(\mathrm{Gd}^{154})$, and $t_\beta(\mathrm{Sm}^{151})$ we find that $t_n(\mathrm{Sm}^{151})\simeq5$ years. Using the ratio of the σ's in this region as compared with iron we find that near iron $t_n\simeq10^3$ years. The mean value of σ near $A=150$ is about 1000 mb whereas $\sigma(\mathrm{Sm}^{152})=1440$ mb. Thus the mean t_n near $A=150$ is of the order of 10 years, and the total time for the capture of ~200 neutrons is about 2×10^3 years. We show later that this shorter of the two time-scales of the *s* process is probably due to the fact that plenty of neutrons are available and that the "steady flow" concept can be used. In this case the σN product remains approximately constant.

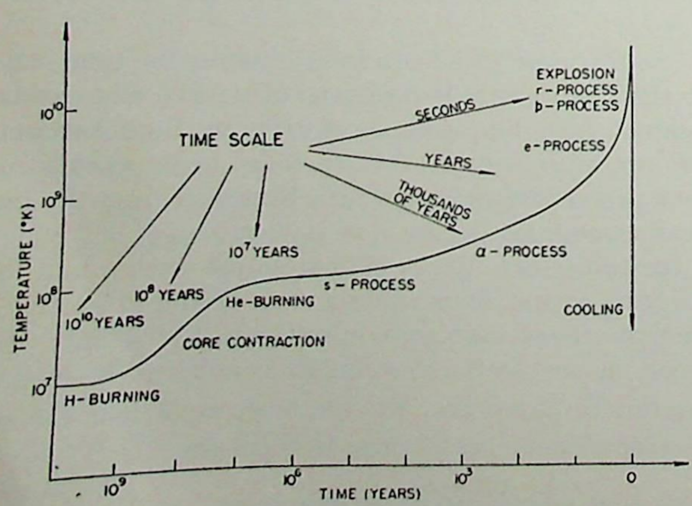


FIG. II,4. The time-scales of the various processes of element synthesis in stars. The curve gives the central temperature as a function of time for a star of about one solar mass. The curve is schematic.

(6) *x-Process Time-Scales*

Several alternative processes which may synthesize deuterium, lithium, beryllium, and boron are described in Sec. X. Depending on which of these is the most satisfactory, the time-scale for synthesis may range from a few seconds if a supernova origin is assumed to $\sim10^9$ years if the surfaces of active stars are considered probable. The time-scale here is important only in as far as it may allow some predictions of the probable distribution of these elements in the solar system and the Galaxy.

In Fig. II,4 we give a schematic diagram showing where the various synthesizing processes take place in a plot of the internal stellar temperature against the

time that the star lives in each temperature range. The estimated time-scales associated with each of these processes are also given very approximately.

III. HYDROGEN BURNING, HELIUM BURNING, THE α PROCESS, AND NEUTRON PRODUCTION

This section and the sections to follow are devoted to detailed elaboration and discussion of the different physical processes introduced in Sec. II. These sections treat quantitatively experimental and theoretical evaluations of the cross sections and reaction rates of the nuclear processes involved in energy generation and element synthesis in stars. The material supplements and extends that published in a series of articles in 1954 (Fo54, Bo54, Ho54), in 1955 (Fo55, Fo55a), and in 1956 (Bu56). In the first part of this section we give a discussion of the relations between nuclear cross sections and nuclear reaction rates in stellar interiors and of the notation used in this and the following sections.

A. Cross-Section Factor and Reaction Rates

The experimental results to be discussed will be used to derive the numerical value of the nuclear cross-section factor for a charged particle reaction defined by

$$S=\sigma(E)E\exp(31.28Z_1Z_0A^{\frac{1}{2}}E^{-\frac{1}{2}})\text{ kev barns}$$

where $\sigma(E)$ is the cross section in barns (10^{-24} cm^2) measured at the center-of-mass energy E in kev. The charges of the interacting particles are Z_1 and Z_0 in units of the proton charge and $A=A_1A_0/(A_1+A_0)$ is their reduced mass in atomic mass units. S is measured in the center-of-mass system. From measurements made in the laboratory system with incident particle energy E_1 and with target nuclei at rest, the quantity S is given by

$$S=\sigma(E_1)E_1\frac{A_0}{A_1+A_0}\exp(31.28Z_1Z_0A_1^{\frac{1}{2}}E_1^{-\frac{1}{2}})\text{ kev barns.}$$

For a nonresonant or off-resonant reaction S is a slowly varying function of the energy E. Methods for extrapolating to the effective thermal energy E_o in stellar interiors have been given by numerous authors (Sa52a, Fo54, Sa55, Ma57). The effective thermal energy at temperature T is

$$E_o=1.220(Z_1^2Z_0^2AT_6^2)^{\frac{1}{3}}\text{ kev barns,}$$

where T_6 is the temperature measured in units of 10^6 °K. The width of the effective range of thermal energy is

$$\Delta E_o=0.75(Z_1^2Z_0^2AT_6^5)^{1/6}\text{ kev.}$$

The mean reaction rate of a thermonuclear process may be expressed as

$$P=\rho r=n_1n_0\langle\sigma v\rangle_{\text{Av}}=3.63\times10^{47}\rho^2\frac{x_1x_0}{A_1A_0}\langle\sigma v\rangle_{\text{Av}}\quad\text{reactions cm}^{-3}\text{ sec}^{-1},$$

where n_1 and n_0 are the number densities of the interacting particles per cm^3 and $\langle\sigma v\rangle_{\text{Av}}$ is the average of the cross section multiplied by the velocity in cm^3 sec^{-1}. The quantity r is the reaction rate per gram per second. The quantities x_1 and x_0 are the amounts of the interacting nuclei expressed as fractions by weight. In terms of $S_o=S(E_o)$ kev barns, it is found for a nonresonant process that

$$P=7.20\times10^{-19}n_1n_0f_oS_o(AZ_1Z_0)^{-1}\tau^2e^{-\tau}\quad\text{reactions cm}^{-3}\text{ sec}^{-1}$$

$$=2.62\times10^{29}\rho^2\frac{x_1x_0}{A_1A_0}f_oS_o(AZ_1Z_0)^{-1}\tau^2e^{-\tau}\quad\text{reactions cm}^{-3}\text{ sec}^{-1},$$

where S_o is in kev barns and

$$\tau=42.48\left(Z_1^2Z_0^2\frac{A}{T_6}\right)^{\frac{1}{3}}$$

(this τ is not to be confused with the mean lifetime of the interacting particles which will always be accompanied by appropriate subscripts, etc.). The term f_o is the electron screening or shielding factor discussed by Salpeter (Sa54), evaluated at E_o. The cross-section factor S_o as customarily calculated does not include allowance for electron screening.

The mean lifetime of the nuclei of type 0 for the interaction with nuclei of type 1 is given by

$$\frac{1}{\tau_1(0)}=p_1(0)=\nu_0P/n_0$$

$$=4.34\times10^5\frac{\rho x_1}{A_1}\nu_0S_of_o(AZ_1Z_0)^{-1}\tau^2e^{-\tau}\text{ sec}^{-1}$$

$$=7.83\times10^8\frac{\rho x_1}{A_1}\nu_0S_of_o\left(\frac{Z_1Z_0}{AT_6^2}\right)^{\frac{1}{3}}e^{-\tau}\text{ sec}^{-1},$$

where ν_0 is the number of nuclei of type 0 consumed in each reaction. The quantity $p_1(0)$ is the mean reaction rate per nucleus of type 0. If nuclei of type 0 are regenerated in a cycle of reactions then $\tau_1(0)$ becomes the mean cycle time for nuclei of type 0.

The most satisfactory procedure for determining S_o is to make experimental observations on cross sections over a range of energies not too large compared to E_o. The cross-section factor, S, can then be plotted as a function of E and an appropriate extrapolation to find S_o can be made. This is not always possible and computational procedures for several frequently occurring cases will now be given.

(i) For the case in which S_o is to be calculated from the experimentally determined parameters of a reso-

nance at E_r which falls outside the range $E_o \pm 2\Delta E_o$, then

$$\sigma = \pi \lambda\!\!\!^{-2} \omega \frac{\Gamma_1 \Gamma_2}{(E-E_r)^2 + \Gamma^2/4},$$

and

$$S_o = 3.10 \times 10^3 \frac{\theta_1^2}{A} \frac{1-\alpha_l E_o}{K_{2l+1}^2(x)} \frac{\omega \Gamma_2 E_R}{(E_o - E_r)^2 + \Gamma^2/4} \text{ kev barns.}$$

The corrections for level shift effects are given by Marion and Fowler (Ma57). The various quantities which enter into these expressions are:

$\lambda\!\!\!^{-}$ = reduced De Broglie wavelength of interacting particles 0 and 1
$= (\hbar^2/2ME)^{\frac{1}{2}} = 144(AE)^{-\frac{1}{2}}$ fermis (10^{-13} cm)
$R/\lambda\!\!\!^{-} = (E/E_R)^{\frac{1}{2}}$
R = interaction radius
$\approx 1.44(A_1^{\frac{1}{3}} + A_0^{\frac{1}{3}})$ fermis
θ_1^2 = dimensionless reduced width for the interaction of particle 1 with particle 0, derivable from the observed width, Γ_1
Γ_1 = width for re-emission of particle 1 with particle 0
$= 6\theta_1^2(\hbar^2/2MR^2)(R/\lambda\!\!\!^{-})P_l$
$= 6\theta_1^2(EE_R)^{\frac{1}{2}}P_l$
$= [3\pi/2K_{2l+1}^2(x)](1-\alpha_l E)\theta_1^2 E_R \times \exp(-31.28 Z_1 Z_0 A^{\frac{1}{2}} E^{-\frac{1}{2}})$
P_l = barrier penetration factor for orbital angular momentum $l\hbar$
$= [\pi(E_R/E)^{\frac{1}{2}}/4K_{2l+1}^2(x)](1-\alpha_l E) \times \exp(-31.28 Z_1 Z_0 A^{\frac{1}{2}} E^{-\frac{1}{2}})$
Γ_2 = experimental width for emission of reaction product 2 with 3; approximately independent of E.
Γ = total width of resonance $= \Gamma_1 + \Gamma_2 + \cdots$
ω = statistical factor $= (2J+1)/(2J_1+1)(2J_0+1)$
J = angular momentum of resonant state
J_1 = angular momentum of particle 1
J_0 = angular momentum of particle 0
$x = 2(E_C/E_R)^{\frac{1}{2}} = 0.525(AZ_1Z_0R)^{\frac{1}{2}}$
$E_R = [1/l(l+1)] \times$ centrifugal barrier height
$= \hbar^2/2MR^2 = 20.9/AR^2 \times 10^3$ kev
E_C = Coulomb barrier height $= Z_1 Z_0 e^2/R$
$= (1.44 Z_1 Z_0/R) \times 10^3$ kev.

$K_{2l+1}(x)$ = modified Bessel function of order $2l+1$. Convenient tables for K are given in (Br37, 52).

$$\alpha_l = \frac{1}{3E_R}\left[\frac{1}{l+1}\left\{\frac{K_{2l+3}(x)}{K_{2l+1}(x)}\left(1 - 2l(l+1)\left(\frac{2}{x}\right)^2\right) - 1\right\} + 2l(l+1)(2l+1)\left(\frac{2}{x}\right)^4\right]$$

$$\alpha_{l=0} = \frac{1}{3E_R}\left\{\frac{K_3(x)}{K_1(x)} - 1\right\}.$$

For large x (high Coulomb barrier):

$$K_{2l+1}(x) \approx (\pi/2x)^{\frac{1}{2}} \exp\left[-x + \frac{2(l+\frac{1}{2})^2}{x}\right]$$

$$P_l \approx \left[\frac{E_C}{E}\right]^{\frac{1}{2}} (1-\alpha_l E) \exp\left[-31.28 Z_1 Z_0 A^{\frac{1}{2}} E^{-\frac{1}{2}} + 4\left(\frac{E_C}{E_R}\right)^{\frac{1}{2}} - 2(l+\tfrac{1}{2})^2 \left(\frac{E_R}{E_C}\right)^{\frac{1}{2}}\right]$$

$$P_l/P_0 \approx \exp[-2l(l+1)(E_R/E_C)^{\frac{1}{2}}] = \exp[-7.62 l(l+1)/(AZ_1Z_0R)^{\frac{1}{2}}]$$

$$\alpha_{l=0} \approx \frac{2}{3(E_R E_C)^{\frac{1}{2}}} \sim 10^{-3} \text{ (kev)}^{-1}.$$

In principle the cross sections and cross-section factors given above must be summed over all values of the orbital angular momentum, $l\hbar$. The expressions given apply to any one specified l-wave interaction. Because of the low energies involved in nuclear collisions in stellar interiors, s-wave ($l=0$) interactions and in some cases p-wave ($l=1$) interactions are usually the only ones which need to be taken into account. However, whenever selection rules or destructive interference effects reduce the low l-value interactions to zero, then higher l values must be considered. In addition, whenever any particular l-wave interaction is resonant, see (iv), then it must be taken into account. Resonant cross sections can be quite large even for high l values.

(ii) For the case in which the cross section is an average over resonances of the same spin and parity with mean width Γ spaced by an average energy interval equal to D, then averaging over the expression for σ given in (i) yields

$$\bar{\sigma} \approx 2\pi^2 \lambda\!\!\!^{-2} (2l+1)\Gamma_1\Gamma_2/D\Gamma.$$

In general, each l wave can form several compound systems of different spins but the same parity. In employing the above approximate relation with the simple statistical factor $(2l+1)$ we assume that these different systems have the same mean widths, Γ, and the same mean level separations, D.

With the foregoing expression for $\bar{\sigma}$, one finds, neglecting $\alpha_l E_o$,

$$S_o \approx 1.94 \times 10^4 \frac{\theta_1^2}{A} \frac{2l+1}{K_{2l+1}^2(x)} \frac{\Gamma_2 E_R}{\Gamma D} \text{ kev barns.}$$

This will prove to be useful for order of magnitude estimates for interactions of charged particles with heavy nuclei ($Z \gtrsim 10$). Blatt and Weisskopf (Bl52) have given a general relation between the particle decay widths, $\Gamma_1 \propto \theta_1^2$, for a given system of resonances and

their mean energy separation. This can be written

$$\theta_1^2 = \frac{\lambda_o}{3\pi R}\frac{D}{E_R}\beta(A_0).$$

The function $\beta(A_0)$ varies between ~ 0.3 and 3 for all except a few light and intermediate weight nuclei (Fe54, We57a) and we will set it equal to unity for order of magnitude estimates in what follows. The quantity $\lambda_o \approx 0.7\times 10^{-13}$ cm is the characteristic reduced De Broglie wavelength of low-energy nucleons once inside nuclei.

We now have, for R in fermis,

$$S_o \sim 1.4\times 10^3 \frac{2l+1}{ARK_{2l+1}^2(x)}\frac{\Gamma_2}{\Gamma} \text{ kev barns.}$$

For $l=0$ and large x we then have

$$S_o(l=0) \sim 1.8\times 10^3 \frac{1}{AR}\frac{\Gamma_2}{\Gamma}\left(\frac{E_C}{E_R}\right)^{\frac{1}{2}} \times \exp\left[+4\left(\frac{E_C}{E_R}\right)^{\frac{1}{2}}\right] \text{ kev barns.}$$

In general for (p,γ) reactions on the light nuclei ($Z=6$ to 12) one finds $S_o \sim 1$ to 100 kev barns. For the (p,α) reactions, $S_o \sim 10^3$ to 10^5 kev barns. On substitution into the expression for the mean lifetime, and setting $\nu_0=1$, $f_o=1$, $l=0$, it is found that

$$\frac{1}{\tau_1(0)} \sim 3.8\times 10^{11} \frac{\rho x_1}{A_1}\frac{(Z_1Z_0)^{5/6}}{R^{\frac{1}{2}}A^{5/6}T_6^{\frac{1}{3}}}\frac{\Gamma_2}{\Gamma} \times \exp\left[+4\left(\frac{E_C}{E_R}\right)^{\frac{1}{2}} - \tau\right] \text{ sec}^{-1}.$$

In cases where this expression is applicable, e.g., (p,γ) or (α,n) reactions on heavy nuclei ($Z \gtrsim 10$) one will in general have $\Gamma \approx \Gamma_2 \gg \Gamma_1$.

(iii) In the case of light nuclei, the interaction energy may fall in the flat minimum in between resonances. For this nonresonant case, order of magnitude estimates may be made by summing the contributions of all resonances of a given type with width Γ spaced by an average energy interval equal to D. The nonresonant cross section for a given l-wave interaction is

$$\sigma_{nr} \approx \pi\lambda^2(2l+1)(\pi^2 \pm 8)\Gamma_1\Gamma_2/D^2.$$

This expression *assumes* that all except the two nearest resonances have random phases, that is, that their contributions to the cross section can be simply added without compounding amplitudes. The factor (π^2+8) is to be used if the two nearest resonances interfere constructively; the factor (π^2-8) if the two nearest resonances interfere destructively. In fact, it is possible for the cross section for a given l wave to go to zero if all of the contributing states interfere to give zero reaction amplitude. In this case the next highest l wave must be considered. On the average, for order of magnitude estimates, we can assume that all resonances, including the two nearest ones, are randomly phased in which case the factor $\pi^2 \pm 8$ becomes π^2. The expression for σ_{nr} is then the same as $\bar{\sigma}$ in (ii) with Γ replaced by $2D/\pi$. With this substitution, expressions for S_o and $1/\tau_1(0)$ given in (ii) can be employed. It will be noted, however, that in general $\Gamma_2 < D$ so that an approximation similar to $\Gamma_2 \sim \Gamma$ is not valid in this case. For the light nuclei $D \sim 1$ to 10 Mev, whereas for radiative processes $\Gamma_2 = \Gamma_\gamma \sim 0.1$ to 100 ev so that $10^{-7} < \Gamma_\gamma/D < 10^{-4}$. For particle emission Γ_2/D may be as high as 0.1.

(iv) When a resonance E_r falls within the range $\sim E_o \pm 2\Delta E_o$ the resonant cross section factor S_r, derivable from the resonant cross section σ_r, must be employed. When resonant, even fairly large l-wave interactions must be considered. We have

$$\sigma_r = 4\pi\lambda^2\omega\Gamma_1\Gamma_2/\Gamma^2.$$

We will consider two limiting cases,

$$\sigma_r \to 4\pi\lambda^2\Gamma_1/\Gamma_2 \quad \text{for} \quad \Gamma \approx \Gamma_2 \gg \Gamma_1,$$
$$\sigma_r \to 4\pi\lambda^2\Gamma_2/\Gamma_1 \quad \text{for} \quad \Gamma \approx \Gamma_1 \gg \Gamma_2.$$

The cross-section factor is

$$S_r = 1.24\times 10^4 \frac{\theta_1^2}{A}\frac{1-\alpha_l E_r}{K_{2l+1}^2(x)}\frac{\omega\Gamma_2 E_R}{\Gamma^2}.$$

The two limiting cases become

$$S_r \to 1.24\times 10^4 \frac{\theta_1^2}{A}\frac{1-\alpha_l E_r}{K_{2l+1}^2(x)}\frac{\omega E_R}{\Gamma_2} \qquad \text{for} \quad \Gamma \approx \Gamma_2 \gg \Gamma_1,$$

$$S_r \to 2.63\times 10^3 \frac{\omega\Gamma_2}{A\Gamma_1} \times \exp(+31.28 Z_1 Z_0 A^{\frac{1}{2}} E_r^{-\frac{1}{2}}) \quad \text{for} \quad \Gamma \approx \Gamma_1 \gg \Gamma_2.$$

The resonant reaction rate is for Γ in kev and S_r in kev barns:

$$P = 3.08\times 10^{-15} n_1 n_0 \frac{f_r}{A^{\frac{1}{2}}T_6^{\frac{3}{2}}} S_r \Gamma e^{-\tau_r} \text{ reactions cm}^{-3}\text{ sec}^{-1}$$

$$= 1.12\times 10^{33}\rho^2 \frac{x_1x_0}{A_1A_0}\frac{f_r}{A^{\frac{1}{2}}T_6^{\frac{3}{2}}} S_r\Gamma e^{-\tau_r}$$

$$\to 1.39\times 10^{37}\rho^2 \frac{x_1x_0}{A_1A_0}\frac{\omega E_R f_r \theta_1^2}{(AT_6)^{\frac{3}{2}}}\frac{1-\alpha_l E_r}{K_{2l+1}^2(x)} e^{-\tau_r} \qquad \text{for} \quad \Gamma \approx \Gamma_2 \gg \Gamma_1$$

$$\to 2.94\times 10^{36}\rho^2 \frac{x_1x_0}{A_1A_0}\frac{\omega\Gamma_2 f_r}{(AT_6)^{\frac{3}{2}}} \exp(-11.61 E_r/T_6) \qquad \text{for} \quad \Gamma \approx \Gamma_1 \gg \Gamma_2.$$

In the preceding expressions

$$\tau_r = 31.28 Z_1 Z_0 A^{\frac{1}{3}} E_r^{-\frac{1}{3}} + 11.61 E_r/T_6$$

and f_r is the electron shielding factor evaluated at E_r.

The mean reaction time of nuclei of type 0 is

$$\frac{1}{\tau_1(0)} = p_1(0) = 1.86\times 10^9 \frac{\rho x_1}{A_1} \frac{\nu_0 f_r}{A^{\frac{1}{3}} T_6^{\frac{2}{3}}} S_r \Gamma e^{-\tau_r}\ \mathrm{sec}^{-1}$$

$$\rightarrow 2.31\times 10^{13} \frac{\rho x_1}{A} \frac{\nu_0 \omega E_R f_r \theta_1^2}{(AT_6)^{\frac{3}{2}}} \frac{1-\alpha_l E_r}{K_{2l+1}^2(x)} e^{-\tau_r}$$

$$\text{for} \quad \Gamma \approx \Gamma_2 \gg \Gamma_1$$

$$\rightarrow 0.49\times 10^{13} \frac{\rho x_1}{A_1} \frac{\nu_0 \omega \Gamma_2 f_r}{(AT_6)^{\frac{3}{2}}} \exp(-11.61 E_r/T_6)$$

$$\text{for} \quad \Gamma \approx \Gamma_1 \gg \Gamma_2.$$

B. Pure Hydrogen Burning

The point of view that element synthesis begins with pure hydrogen (primordial or continuously produced) condensed in stars is based on the existence of the so-called direct *pp* chain of reactions by which hydrogen is converted into helium. This chain is initiated by the direct *pp* reaction

$$p + p \rightarrow d + \beta^+ + \nu_+ + 0.421\ \text{Mev},$$

which has good theoretical foundations but which has not yet been observed experimentally in the laboratory because of its extremely low cross section even at relatively high interaction energies. In the above equation and in what follows we use ν_+ for neutrinos emitted with positrons, β^+, and ν_- for antineutrinos emitted with electrons, β^-. *We use nuclear rather than atomic mass differences in expressing the Q values of all reactions.* There is of course practically no difference in atomic and nuclear Q values when positrons or electrons are not involved.

The calculated cross section for the *pp* reaction is 10^{-47} cm^2$=10^{-23}$ barn at 1-Mev laboratory energy. This is much too small for detection with currently available techniques. The yield when a thick hydrogen target is employed can be estimated to be 5×10^{-28} reaction per incident proton or approximately one reaction per 3×10^8 amp sec corresponding at 10^6 volts to 3×10^{14} w sec. Thus only one reaction per *10 Megawatt years* of bombardment can be expected. In an earlier paper (Fo54) it was indicated that the cross section would reach a maximum value at 100 kev. This is not the case if account is taken of the rapid increase in beta-decay probability with increasing beta-decay energy and hence with increasing incident proton energy. At the present time, however, it is not possible to obtain within many orders of magnitude, the necessary amounts of bombardment even at 1 Mev or above. At 100 Mev the cross section becomes 3×10^{-15} barn but very pure hydrogen targets would be necessary to avoid positron production in heavy nuclei as contaminants. At still higher energy, meson production becomes a complication.

(1) Rate of pp Reaction

Definitive theoretical calculations on the rate of the *pp* processes were given by Salpeter (Sa52a). These calculations accurately took into account the well-determined phenomenological parameters of the neutron-proton and proton-proton interactions and showed that the main uncertainty in the cross section lay at that time in the uncertainty of the Gamow-Teller beta-decay constant for which the following value was given:

$$g = 7.5\pm 1.5\times 10^{-4}\ \mathrm{sec}^{-1}\ \text{(Sa52a)}.$$

This value was adopted by Fowler (Fo54). The 20% uncertainty in the value of the Gamow-Teller coupling constant did not imply any uncertainty in the validity of the Gamow-Teller selection rules for beta-decay transitions. Then, as now, the super-allowed nature of the beta decay, $He^6(0^+)\rightarrow Li^6(1^+)+\beta^-+\nu_-$, which is permitted only by the Gamow-Teller selection rules, serves as a solid empirical foundation for this type of transition. The recent discovery of the nonconservation of parity‡ in allowed beta decay transitions (Wu57, Le56, Al57) is further substantiation of the existence of Gamow-Teller selection rules. The fore-and-aft asymmetry in the decay of aligned Co^{60} nuclei observed by Wu *et al.* is not possible for allowed Fermi transitions even with nonconservation of parity.

Experimental data on *ft* values for super-allowed mirror transitions and $0\rightarrow 0$ transitions in beta decay available in 1956 were employed by Kofoed-Hansen and Winther (Ko56) to evaluate the coupling constants in beta decay with improved accuracy. A recent measurement of the energy of the $O^{14}(0^+)\rightarrow N^{14*}(0^+)$ transition by Bromley *et al.* (Br57) leads to only a minor change in Kofoed-Hansen and Winther's results. Bromley *et al.* find 1809.7 ± 7.8 kev for the transition energy and thus an *ft* value equal to 3088 ± 56, which brings O^{14} into agreement with the Al^{26}, Cl^{34}, and K^{38} $0^+\rightarrow 0^+$ transitions used by Kofoed-Hansen and Winther.

The quantity g which appears in Salpeter's expression for the rate of the *pp* reaction is related to the quantities B and x defined by Kofoed-Hansen and Winther (Ko56) through the relation

$$g = 4\ln 2(x/B) = 2.77(x/B).$$

They find that the best fit to the experimental data yields $B=2787\pm 70$ sec, $x=0.560\pm 0.012$, so that $B/x=4980\pm 150$ sec. With the new O^{14} data we estimate $B=2818$, $x=0.544$ so that $B/x=5180$ and

$$g = 5.35\pm 0.15\times 10^{-4}\ \mathrm{sec}^{-1}\ \text{(new value, 1957)},$$

‡ The reader is admonished that parity is still conserved to a high degree of experimental accuracy in reactions involving the strong interactions among nucleons (Ta57) as contrasted with the weak interactions leading to beta decay. The importance of parity conservation in nuclear reactions is illustrated in the discussion of the 3α reaction in Sec. IC(1).

TABLE III,1. Energy generation in pp chain and CN cycle.

	pp chain		CN cycle			
			$N^{14}(p,\gamma)$ determining		$C^{12}(p,\gamma)$ determining	
T 10^6 °K	E_o kev	$\text{Log}\left(\frac{100}{\rho x_H^2 f_{pp}}\epsilon_{pp}\right)$	E_o kev	$\text{Log}\left(\frac{1}{\rho x_H x_N f_N}\epsilon_N\right)$	E_o kev	$\text{Log}\left(\frac{1}{\rho x_H x_C f_C}\epsilon_C\right)$
5	2.8	−0.79	13	−11.24	11	−7.71
10	4.5	+0.79	20	−3.47	18	−0.73
15	5.9	1.54	27	+0.30	24	+2.65
20	7.1	2.01	32	2.67	29	4.78
30	9.3	2.58	42	5.63	38	7.45
40	11.3	2.94	51	7.49	46	9.13
50	13.2	3.19	60	8.81	53	10.32
70	16.5	3.54	75	10.72	67	12.0
100	20.8	3.86	94	12.3	84	13.5

which is 29% less than the value given by Salpeter (Sa52a) and used by Bosman-Crespin *et al.* (Bo54) in numerical calculations. These numerical values should thus be reduced by 29%. The over-all probable error, taking other sources of error into account, can be conservatively estimated as 10%. The cross-section factor becomes

$$S = 28.5(1+0.008E)\times10^{-23} \text{ kev barn } \pm 10\%$$

for E in kev or

$$S_o = 28.5(1+0.008T_6^{\frac{1}{3}})\times10^{-23} \text{ kev barn } \pm10\%.$$

The correction term in E or $T_6^{\frac{1}{3}}$ arises from the dependence on energy of the beta-decay transition probability, and in addition from small corrections to the barrier penetrability when a nonzero interaction radius is assumed. It is accurate only for $E<100$ kev or $T<10^9$ degrees. At laboratory bombarding energies for $E\gtrsim1$ Mev, $S\sim2E^5\times10^{-24}$ Mev barn, $\sigma\sim3E_o^{4.5}\times10^{-24}$ barn. In the following discussion it is convenient to use a quantity α_{pp}, which is involved in the reaction rate and is given by

$$\alpha_{pp} = \rho f_{pp} T_6^{-\frac{2}{3}}(1+0.012T_6^{\frac{1}{3}}+0.008T_6^{\frac{2}{3}}+0.00065T_6)$$
$$\times\exp(-33.804T_6^{-\frac{1}{3}})\pm10\%$$
$$\simeq 2.0\times10^{-7}\rho f_{pp}\left(\frac{T_6}{15}\right)^{3.95} \quad \text{near} \quad T_6=15.$$

The electron screening factor, f_{pp}, can be taken as approximately equal to unity for $T_6\gtrsim8$. The quantity ρ is the density and x_H is the concentration by mass of hydrogen. The correction terms in $T_6^{\frac{1}{3}}$ and T_6 arise in integrating over all the interaction energies. The first has been discussed by Salpeter (Sa52a) and the second comes when the energy dependence of S is included to first order in the integration. The reaction rate in reactions per gram of material per second is $\alpha_{pp}x_H^2$ multiplied by a numerical factor derivable from S_o, *viz.*,

$$r_{pp} = 0.833\alpha_{pp}x_H^2\times10^{11} \text{ reactions g}^{-1}\text{ sec}^{-1}$$

while the mean reaction rate per proton with *two protons* consumed per reaction is

$$p_{pp}(\text{H}) = -\frac{1}{x_H}\frac{dx_H}{dt} = 2.78(\alpha_{pp}x_H)\times10^{-13}\text{ sec}^{-1}.$$

If the pp chain is completed then the energy release in the over-all process, $4H^1\rightarrow He^4$, is 26.22 Mev = 4.201 $\times10^{-5}$ erg, using the atomic masses of Wapstra (Wa55). This value excludes the 2% neutrino energy loss. The expression for the energy generation, ϵ_{pp}, is then given by

$$\epsilon_{pp} = 0.5\times4.201\times10^{-5}r_{pp} \text{ ergs}$$

or

$$\epsilon_{pp} = 1.75\alpha_{pp}x_H^2\times10^6 \text{ ergs g}^{-1}\text{ sec}^{-1} \quad (4H^1\rightarrow He^4).$$

This energy generation as a function of temperature is tabulated in Table III,1 and is indicated graphically in

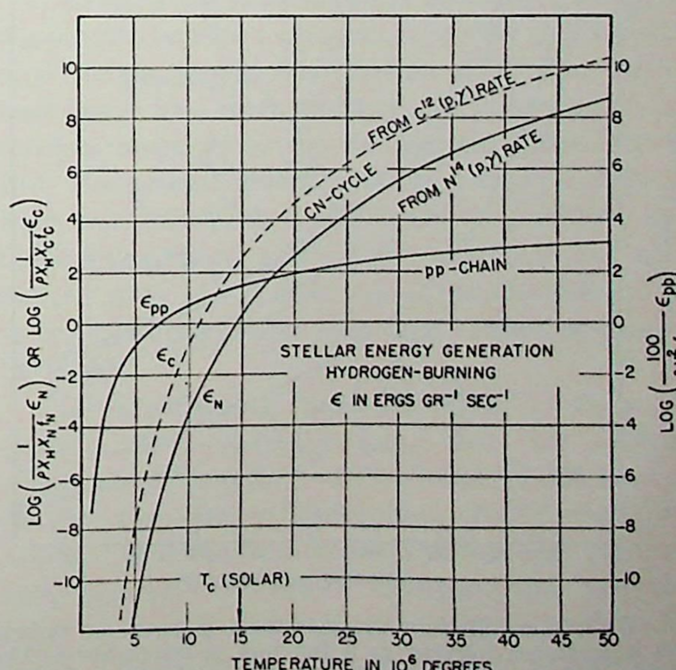


FIG. III,1. Energy generation in ergs g^{-1} sec^{-1} as a function of temperature in the pp chain and the CN cycle. The ordinates give $\log\epsilon$ directly for $\rho x_H^2 f_{pp}=100$ g/cm³, $\rho x_H x_N f_N=1$ g/cm³ or $\rho x_H x_C f_C=1$ g/cm³. They are thus appropriate for $x_N=x(N^{14})$ or $x_C=x(C^{12})\simeq0.8x(C^{12}+C^{13})=0.01x_H$. If $N^{14}(p,\gamma)$ is resonant the CN-cycle rate is determined by $C^{12}(p,\gamma)$ and $x_C\simeq0.8x_{CN}$. If $N^{14}(p,\gamma)$ is nonresonant, it determines the CN-cycle rate and $x_N\approx x_{CN}$.

TABLE III,2. Mean lifetimes or cycle times in hydrogen burning (years).[a]

T 10^6 °K	$\text{Log}\rho x_C f_C \tau(\text{H})$ H^1 $\text{C}^{12}(p,\gamma)$	$\text{Log}\rho x_N f_N \tau(\text{H})$ H^1 $\text{N}^{14}(p,\gamma)$	$\text{Log}(\rho x_H f\tau/100)$ H^1 pp chain	C^{12}	N^{14}	O^{16}	Ne^{20}	Ne^{21}	Ne^{22}	T (10^6 °K)
5	+19.0	+22.5	+12.09	+16.5	20.0					5
10	12.0	14.7	10.51	9.53	12.2					10
15	8.63	11.0	9.76	6.15	8.44	10.8				15
20	6.50	8.61	9.29	4.02	6.07	8.24	+10.92	+12.41	+10.62	20
30	3.83	5.65	8.72	1.35	3.11	4.99	7.28	8.63	6.76	30
40	2.15	3.79	8.36	−0.33	1.25	2.94	5.02	6.25	4.32	40
50	0.96	2.47	8.11	−1.52	−0.07	1.49	3.42	4.55	2.55	50
70	−0.72	0.56	7.76	−3.2	−1.98	−0.50	1.20	2.18	0.05	70
100	−2.22	−1.00	7.44	−4.7	−3.54	−2.37	−0.80	0.02	−2.33	100

[a] O^{16} and Ne^{20} lifetimes based on results of Ta57a and Pi57a.

Fig. III,1, where it is compared with that for the CN cycle, which is discussed in Sec. III F.

If the chain goes only to He^3, as might be the case in limited times at low temperature, then the energy release in $3\text{H}^1 \rightarrow \text{He}^3$ is 6.68 Mev $= 1.070 \times 10^{-5}$ erg and

$$\epsilon_{pp}' = 0.89 \alpha_{pp} \times 10^6 \text{ ergs g}^{-1}\text{ sec}^{-1} \quad (3\text{H}^1 \rightarrow \text{He}^3).$$

The mean lifetime of hydrogen in the pp chain becomes

$$\tau_{pp}(\text{H}) = 1.14(\alpha_{pp} x_\text{H})^{-1} \times 10^5 \text{ years}$$

$$= 1.99(x_\text{H}/\epsilon_{pp}) \times 10^{11} \text{ years}$$

$$= \frac{1.99 \times 10^9}{E_{pp}} \frac{100}{\rho x_\text{H}} \text{ years},$$

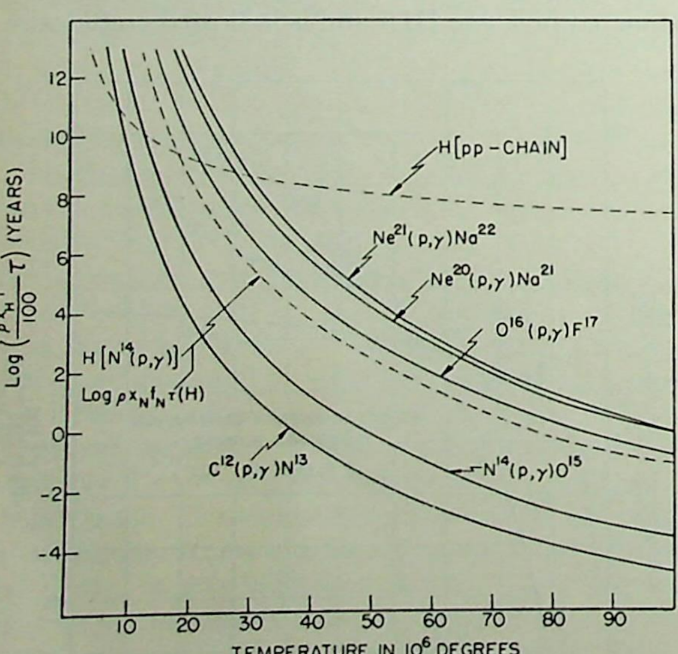


FIG. III,2. Mean lifetimes or cycle times in years for various nuclei in hydrogen burning as a function of temperature. The ordinate gives $\log\tau$ (years) directly for $\rho x_\text{H} f = 100$ g/cm³ or $\rho x_N f_N = 1$ g/cm³. The hydrogen lifetime in the CN cycle is calculated from the rate of the nonresonant $\text{N}^{14}(p,\gamma)$ reaction. The mean reaction time of the pp reaction is twice that of the hydrogen since two protons are consumed in each reaction. The latest experimental results (Ta57a) on $\text{O}^{16}(p,\gamma)$ require that the curve for $\tau(\text{O}^{16})$ be lowered by 1.6 in the logarithm. Those on $\text{Ne}^{20}(p,\gamma)$ require that $\tau(\text{Ne}^{20})$ be lowered by 0.8 in the logarithm (Pi57a).

where E_{pp} is the quantity tabulated by Bosman-Crespin *et al.* (Bo54) decreased by 29%. It will be recalled that $\rho x_\text{H} \sim 100$ g/cm³ at the center of the sun and similar stars. Hydrogen lifetimes at relevant temperatures are given in Table III,2 and Fig. III,2.

The recent detection (Re53, Co56a) of the free antineutrino through absorption by hydrogen in the reaction

$$\nu_- + p \rightarrow \beta^+ + n - 1.804 \text{ Mev}$$

has led to increased confidence in Fermi's neutrino theory of beta decay, and thus in the theoretical calculations of the rate of the pp reaction which have just been discussed. In addition, the interesting possibility arises of investigating antineutrino capture in deuterium in processes which would be inverse to the direct pp reaction. Since antineutrinos are produced in the negative beta decay in a neutron reactor, observation of the following reaction

$$\nu_+ + d \rightarrow p + p + \beta^- - 1.443 \text{ Mev},$$

which is directly inverse to the pp reaction, would require that antineutrinos and neutrinos be identical. Recent experiments (Aw56), which show that the half-lives of Ca^{48} and Zr^{96} to double beta decay are $\gtrsim 10^{18}$ years, indicate that this is not the case. In addition, if $\nu_- \neq \nu_+$, the reaction $\text{Cl}^{37} + \nu_- \rightarrow A^{37} + \beta^-$ would not be expected to take place and this has been shown to be the case (Da56). Also, the $\nu_+ + d$ process would be very difficult to detect over background because the energy given to the protons and to the electron will be dissipated promptly in electronic collisions. The process will not have a distinguishing signature as does the antineutrino absorption by protons, where the slowing down of the neutron gives an observable delay up to as much as 17 μsec (Co56a) between the positron annihilation energy release and the neutron capture energy release.

On the other hand, the "mirror-inverse" process to the pp reaction, which is

$$\nu_- + d \rightarrow n + n + \beta^+ - 4.029 \text{ Mev},$$

can be induced by reactor antineutrinos of sufficient energy and it will have a signature even more distinctive than that for the absorption in light hydrogen. Only a

small fraction of the reactor antineutrinos have energies above the threshold for this process and thus it will be much more difficult to observe than the absorption by protons. In fact, in the current experiments by Cowan *et al.* (Co56a) deuterium is used in place of part of the hydrogen in order to show that the observed events are directly proportional to the proton concentration. While this paper has been in preparation, two papers have appeared on the antineutrino disintegration of the deuteron (Mu57, We57) which give the effective cross sections for reactor neutrinos as $\bar{\sigma}(\nu_- + p) = 6\times10^{-44}$ cm^2 and $\bar{\sigma}(\nu_- + d) = 2\times10^{-45}$ cm^2. These must be multiplied by a factor of 2 in order to account for parity-nonconserving beta transitions. In any case the deuteron disintegration cross section is only 1/30 of that for the proton. However, if the mirror inverse process can ultimately be observed, the observation would constitute fundamental experimental confirmation of the very general ideas underlying the *pp* process. *Note added in proof.*—Cowan and Reines (Co57a) have not been successful in detecting the $\nu_- + d$ reaction but their experimental upper limit of $\bar{\sigma}(\nu_- + d) = 4\times10^{-45}$ cm^2 does not exclude the theoretically expected value given above.

(2) *Other Reactions in the pp Chain*

New measurements of the cross section for $D^2(p,\gamma)$ have been reported (Gr55) since the review by Fowler (Fo54), but these new values do not necessitate any significant change in the value of the cross-section factor given there ($S_o = 8\times10^{-5}$ kev barn). The final results for the $He^3(He^3,2p)He^4$ cross section have been published (Go54). The value of 2.5×10^{-6} barn at $E_{lab} = 200$ kev or $E_{cm} = 100$ kev yields a cross section factor equal to 1200 kev barn. Fairly substantial corrections for nonzero radius of interaction in the barrier penetrability energy dependence and for electron shielding effects will increase this to an effective value, $S_o = 2000$ kev barns, at relevant stellar energies.

C. Pure Helium Burning

When hydrogen burning in a star's main-sequence stage leads eventually to hydrogen exhaustion, a helium core remains at the star's center. It has been suggested (Sa52, Sa53, Op51, Op54) that the fusion of helium plays an important role in energy generation and element synthesis in the red-giant stage of the star's evolution. The fusion occurs through the processes

$$He^4 + He^4 \rightleftharpoons Be^8$$

$$Be^8 + He^4 \rightleftharpoons C^{12*} \rightarrow C^{12} + \gamma$$

or, in a more condensed notation, through

$$3He^4 \rightleftharpoons C^{12*} \rightarrow C^{12} + \gamma.$$

We refer to this as the 3α reaction. These processes are believed to occur at a late stage of the red giant evolution in which the hydrogen in the central core has been largely converted into helium, and in which gravitational contraction (Ho55) has raised the central temperature to $\sim10^8$ degrees, and the density to $\sim10^5$ g/cc. Under these conditions, as shown by Salpeter, an equilibrium ratio of Be^8 to He^4 nuclei equal to $\sim10^{-9}$ is established. This conclusion followed from experimental measurements (He48, He49, To49, Wh41) which established the fact that Be^8 was unstable to disintegration into two alpha particles but only by 95 kev with an uncertainty of about 5 kev.

(1) *Rate of the 3α Reaction*

Even though very small, the equilibrium concentration of Be^8 is sufficient to lead to considerable production of C^{12} through radiative alpha-particle capture by the Be^8, and of O^{16}, Ne^{20}, etc., by succeeding alpha-particle captures. Salpeter's calculations, in which resonance effects due to the Be^8 ground state were taken into account, indicated a rate for the helium-burning considerably greater at a given pressure and temperature than that previously calculated by Öpik using nonresonant reaction rates (Op54). Detailed consideration of the reaction rates and of the resulting relative abundances of He^4, C^{12}, and O^{16} led Hoyle (Ho54) to the prediction that the foregoing second reaction, in which C^{12} is produced, must exhibit resonance within the range of energies at which the interaction between Be^8 and He^4 effectively occurs. Hoyle's predicted value for the resonance energy was 0.33 Mev, corresponding to an excited state in C^{12} at 7.70 Mev.

It recently has been shown (Fo56, Co57) through the studies of the alpha emission following the beta decay of B^{12} to the second excited state of C^{12}, that this state does decay into three alpha particles through an intermediate stage involving He^4 and the ground state of Be^8. The general reversibility of nuclear reactions thus leads one to expect that C^{12} will be produced in the Salpeter reactions given above and that the second excited state of C^{12} does indeed play a dominant role in the synthesis of elements from helium, as predicted by Hoyle.

The experiments reported (Co57) show that $Q(Be^8 - 2He^4) = 93.7\pm0.9$ kev, that the excitation energy of C^{12*} is $Q(C^{12*} - C^{12}) = 7.653\pm0.008$ Mev, that $Q(C^{12*} - Be^8 - He^4) = 278\pm4$ kev, $Q(C^{12*} - 3He^4) = 372 \pm4$ kev, and that the most probable spin and parity assignments for C^{12*} are 0^+. The state must have either even spin with even parity or odd spin with odd parity. Otherwise, by the conservation of angular momentum and parity, it cannot decay into three alpha particles through the ground state of Be^8. Since the Be^8 is in equilibrium, the reaction rate for the conversion of helium into carbon can be calculated simply from the $3He^4 \rightleftharpoons C^{12*}$ equilibrium with a small leakage to

stable C^{12}. The reaction rate per alpha particle is

$$p_{3\alpha}(He^4)=\frac{1}{\tau_{3\alpha}(He^4)}=\frac{1}{x_\alpha}\frac{dx_\alpha}{dt}$$

$$=8\pi^3 3^{\frac{3}{2}}\frac{\hbar^5}{M_\alpha{}^5(kT)^3}(\rho x_\alpha)^2\frac{\Gamma_\gamma\Gamma_\alpha}{\Gamma_\gamma+\Gamma_\alpha}\times\exp(-Q/kT)\ \text{sec}^{-1},$$

where the appropriate Q is $Q(C^{12*}-3He^4)=372\pm4$ kev. Numerically:

$$p_{3\alpha}(He^4)=\frac{1}{\tau_{3\alpha}(He^4)}=2.37\times10^{-4}(\rho x_\alpha)^2\frac{\Gamma_\gamma}{T_8{}^3}\exp\left(-\frac{43.2}{T_8}\right)\ \text{sec}^{-1},$$

where Γ_γ is in ev and T_8 is the temperature in units of 10^8 degrees. Electron screening will increase this rate by a small factor. Using Salpeter's expression for weak screening (Sa54), we find that $f_{3\alpha}\simeq\exp 0.88(\rho_5/T_8{}^3)^{\frac{1}{2}}$, where ρ_5 is the density in 10^5 g/cm³. Kavanagh (Ka56) finds that $\Gamma_\alpha\gg\Gamma_\gamma$ for C^{12*}, so that Γ_α does not appreciably influence the reaction rate numerically. Farrell (Fa57), Salpeter (Sa57), and Hayakawa *et al.* (Ha56b) have calculated Γ_γ on a reasonable model for states of C^{12} and find $\Gamma_\gamma\sim0.001$ ev. If the γ radiation is highly forbidden then Γ_γ must be replaced by the width for pair emission by C^{12*}. From the inelastic scattering (Fr55) of high-energy electrons by C^{12} this can be calculated (Fr56, Op39) to be $\Gamma_{e\pm}\sim5\times10^{-5}$ ev.

The present rate is in reasonable agreement with that given by Hoyle (Ho54), although it is worth pointing out that two changes have been made from these earlier calculations. Formerly Γ_γ was taken as 1 ev, thereby exaggerating the reaction rate. On the other hand,

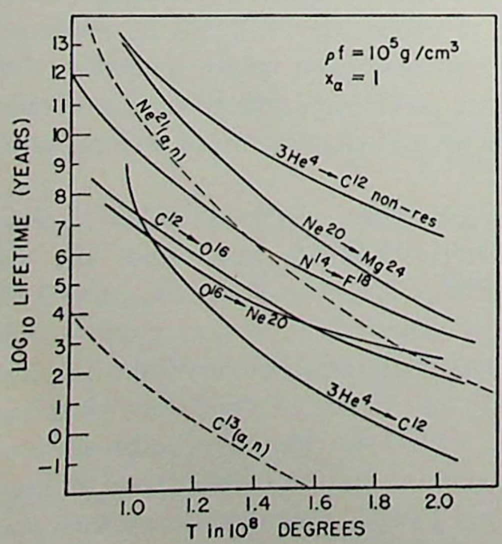


FIG. III,3. Mean lifetimes in years for various nuclei in helium burning as a function of temperature. The curves are drawn in each case for $x_\alpha=1$ and $\rho f=10^5$ g/cm³. The He⁴ lifetime to $3He^4\to C^{12}$ is inversely proportional to $(\rho x_\alpha)^2 f_{3\alpha}$. The other lifetimes are inversely proportional to $\rho x_\alpha f$.

the resonance level was formerly taken such that $Q(C^{12*}-3He^4)=420$ kev. This reduced the reaction rate by an amount that closely compensated for the error in Γ_γ.

The energy generation, with $Q(3He^4-C^{12})=7.281$ Mev $=1.165\times10^{-5}$ erg, becomes

$$\epsilon_{3\alpha}=1.39\times10^{14}(\rho^2x_\alpha{}^3)f_{3\alpha}\frac{\Gamma_\gamma}{T_8{}^3}\exp\left(-\frac{43.2}{T_8}\right)\ \text{ergs g}^{-1}\ \text{sec}^{-1}.$$

If the C^{12} formation is followed by rapid production of O^{16} and Ne^{20}, then $Q(5He^4-Ne^{20})=19.18$ Mev $=3.072\times10^{-5}$ erg and the numerical coefficient in the above equation becomes 3.66×10^{14}. Salpeter (Sa57) and Hayakawa *et al.* (Ha56b) have shown that this rate of energy generation is ample to supply the energy emitted by a star while it is on the tip of the red-giant branch when $\rho\sim10^5$ g/cm³, $x_\alpha\sim1$, and $T\sim10^8$ degrees. Under these conditions $\epsilon_{3\alpha}\approx600$ ergs g⁻¹ sec⁻¹ and $\epsilon_{5\alpha}\approx1500$ ergs g⁻¹ sec⁻¹.

(2) *Rate of the $C^{12}(\alpha,\gamma)O^{16}$ Reaction*

The C^{12} produced in the helium fusion can capture an alpha particle to form O^{16} and thus continue the element synthesis. The binding energy for an alpha particle in O^{16} is 7.148 Mev, so that this process proceeds mainly through the bound state in O^{16} at 7.116 Mev. This state has spin and parity equal to 1^- and thus the alpha capture by $C^{12}(0^+)$ is a p-wave capture and the radiation to the ground state of $O^{16}(0^+)$ is electric dipole. The probability for this radiation is somewhat reduced by the isotopic spin selection rule for $T=0\to0$ for electric dipole transitions. Swann and Metzger (Sw56) find the mean lifetime of the state to be $0.5_{-0.2}{}^{+0.8}\times10^{-14}$ sec, or $\Gamma_\gamma\approx0.13$ ev. The reduced alpha particle width for the state can be taken as $\theta_\alpha{}^2=0.1$ with a probable variation of a factor of 10 either way (Aj55). The cross-section factor becomes

$$S_o=\frac{3.8\times10^7}{(E_o+32)^2}\ \text{kev barns}$$

for $E_o=199T_8{}^{\frac{2}{3}}$ kev. The reaction rate per C^{12} nucleus is

$$\frac{1}{\tau_\alpha(C^{12})}=p_\alpha(C^{12})=-\frac{1}{x_{12}}\frac{dx_{12}}{dt}$$

$$=1.44\times10^7\rho x_\alpha S_oT_8{}^{-\frac{2}{3}}\exp(-69.19T_8{}^{-\frac{1}{3}})\ \text{sec}^{-1}$$

$$=1.38\times10^{10}\frac{\rho x_\alpha}{T_8{}^2(1+0.16T_8{}^{-\frac{1}{3}})^2}\times\exp(-69.19T_8{}^{-\frac{1}{3}})\ \text{sec}^{-1}.$$

If a screening term is introduced it will be the same as $f_{3\alpha}$ in the 3α reaction.

(3) *Further Alpha-Capture Reactions*

The rate of the $O^{16}(\alpha,\gamma)Ne^{20}$ process depends critically on whether the excited states in Ne^{20} at 4.95 and 5.62 Mev have the proper spin and parity (even-even or odd-odd) to be formed by O^{16} and the alpha particles in their ground states. These states probably have even parity and even spin and since they occur at 204 kev and 874 kev, respectively, in $O^{16}+\alpha$ and since $E_o=246T_8^{\frac{2}{3}}$ kev it is clear that for $T\gtrsim 10^8$ degrees they will act as resonances in the conversion of O^{16} into Ne^{20}. Production of Ne^{20} in comparable or greater amounts than O^{16} in helium burning is thus to be expected. Resonances can also be expected to occur in $Ne^{20}(\alpha,\gamma)Mg^{24}$ but the Mg^{24} production will be small because of the large barrier factor for alpha particles.

In Fig. III,3 and Table III,3 we show the mean lifetimes *versus* temperature for the $3He^4\rightarrow C^{12}$ and $C^{12}(\alpha,\gamma)O^{16}$ reactions as calculated above, as well as the lifetimes for $O^{16}(\alpha,\gamma)Ne^{20}$ and $Ne^{20}(\alpha,\gamma)Mg^{24}$ as calculated by Salpeter (Sa57). Salpeter uses a dimensionless alpha-particle width, $\theta_\alpha^2=0.02$, for $O^{16}(\alpha,\gamma)$. We have taken $\theta_\alpha^2=0.01$ for $Ne^{20}(\alpha,\gamma)$ and have modified his calculations accordingly. We also show in Fig. III,3 the nonresonant $3He^4\rightarrow C^{12}$ lifetime to illustrate the large factor arising from the resonance predicted by Hoyle.

In comparing the rates of the helium-burning reactions with the atomic abundances of C^{12}, O^{16}, Ne^{20}, and Mg^{24} two points must be borne in mind. In the first place only the rate of the $3He^4\rightarrow C^{12}$ is known from experimental evidence to better than a factor of ten. The others are very uncertain and one of the most pressing current problems in element synthesis is the need for experimental measurements on $C^{12}(\alpha,\gamma)$, $O^{16}(\alpha,\gamma)$, and $Ne^{20}(\alpha,\gamma)$. Measurements at high energies (2–5 Mev), where the reactions can be easily detected, are needed to serve as guides in extrapolating to lower stellar energies (0.2–0.5 Mev). Such measurements are planned in the near future at the Kellogg Radiation Laboratory.

In the second place, the relative rates of the reactions depend critically on temperature, and thus different astrophysical circumstances will lead to marked differences in abundance ratios. Helium burning can result, for example, in large or small C^{12}/Ne^{20} ratios. At the same time it is clear from Fig. III,3 that in the region 1.0 to 1.3×10^8 degrees the reaction rates are all of the same order of magnitude within the present uncertainties except for the very slow production of Mg^{24} from Ne^{20}. Thus the atomic abundance ratios of $C^{12}:O^{16}:Ne^{20}:Mg^{24}=1:6:2:0.2$, given by Suess and Urey (Su56), or 1:5:8:1, given by Alle (Al57a) for early-type stars, are reasonable value on the basis of the synthesis of these elements in helium burning. Also other processes are involved. A considerable amount of C^{12} has been converted into N^{14} by the CN cycle (see Secs. III F and XI A); Mg has been produced in the α process (see Sec. III D which consumes O^{16} and Ne^{20}, and so on.

D. α Process

As has just been discussed, helium-burning synthesizes C^{12}, O^{16}, Ne^{20}, and perhaps a little Mg^{24}. This occurs at temperatures between 10^8 and 2×10^8 degree and results in the exhaustion of the helium produced in hydrogen-burning. An inner core of C^{12}, O^{16}, and Ne develops in the star and eventually undergoes gravitational contraction and heating just as occurred previously in the case of the helium core. Calculations of stellar evolutionary tracks have not yet been carrie to this stage, but it is a reasonable extrapolation of current ideas concerning the cause of evolution int the giant stage. Gravitation is a "built-in" mechanism in stars which leads to the development of high temperature in the ashes of exhausted nuclear fuel. Gravitatio takes over whenever nuclear generation stops; it raise the temperature to the point where the ashes of th previous processes begin to burn. Implicit in th argument is the assumption that mixing of core an surrounding zones does not occur.

No important reactions occur among C^{12}, O^{16}, an Ne^{20} until significantly higher temperatures, of th order of 10^9 degrees, are attained. Two effects the arise. The γ rays present in the thermal assemb become energetic enough to promote $Ne^{20}(\gamma,\alpha)O$ This is the first (γ,α) reaction to occur since the alph particle binding in Ne^{20} is only 4.75 Mev, while it 7.37 Mev in C^{12} and 7.15 Mev in O^{16}. The proton an neutron binding energies in these nuclei are 12 to Mev, so that (γ,α) precedes (γ,p) or (γ,n). The alp

TABLE III,3. Mean lifetimes in helium burning (years).

T 10^6 °K	$\text{Log}\left[\frac{\rho x_\alpha}{10^5}\right]^2 f_{3\alpha}\tau(He^4)$ $3He^4\rightarrow C^{12}$	$\text{Log}\frac{\rho x_\alpha f}{10^5}\tau$ $C^{12}(\alpha,\gamma)$	$N^{14}(\alpha,\gamma)$	$O^{16}(\alpha,\gamma)$	$Ne^{20}(\alpha,\gamma)$	$C^{13}(\alpha,n)$	$Ne^{21}(\alpha,n)$
70	+15.47	+11.1	+13.6	11.6	17.9	+5.44	+16.3
100	7.90	7.57	9.60	7.5	12.6	1.93	10.86
120	5.01	5.94	7.78	5.9	10.1	0.31	8.31
150	2.18	4.09	5.70	4.4	7.3	−1.54	5.41
175	0.59	2.90	4.36	3.5	5.5	−2.73	3.59
200	−0.58	1.91	3.26	2.9	4.0	−3.71	2.03
250	−2.16	0.39	1.54	2.0	1.7	−5.23	−0.38
300	−3.18	−0.77	0.23	1.5	−0.1	−6.39	−2.20

particles thus released can now penetrate the Coulomb barrier of the remaining Ne^{20} quite readily at 10^9 degrees, thereby forming Mg^{24}. The resulting transformation is strongly exothermic, *viz.*:

$$Ne^{20}+\gamma\rightarrow O^{16}+He^4-4.75\text{ Mev}$$

$$Ne^{20}+He^4\rightarrow Mg^{24}+\gamma+9.31\text{ Mev.}$$

Combining these two equations yields:

$$2Ne^{20}\rightarrow O^{16}+Mg^{24}+4.56\text{ Mev.}$$

Of course, a proportion of the released alpha particles is consumed in scouring out the C^{12} and forming O^{16}. The result is the formation of additional O^{16} and Mg^{24}; this is highly favored on a statistical basis even though complete equilibrium is not attained.

The addition of the alpha particles released from the Ne^{20} need not be limited to the formation of Mg^{24}. Thus, once some Mg^{24} is produced we also expect $Mg^{24}(\alpha,\gamma)Si^{28}$ to take place, since it is also possible to penetrate the Coulomb barrier of Mg^{24} at temperatures of order 10^9 degrees. Again, once an appreciable concentration of Si^{28} is built up, we expect $Si^{28}(\alpha,\gamma)S^{32}$ to take place in some degree, and so on for the production of A^{36} and Ca^{40}.

This is the α process marked in Figs. I,1 and I,2, and responsible, in our view, for building, in decreasing proportion, the α-particle nuclei Mg^{24}, Si^{28}, S^{32}, A^{36}, and Ca^{40}. Their relative atomic abundances are indicated in Table III,4. The abundance of Ne^{20} is included for comparison. The proportions decrease in general along this sequence, partly because the production of any member of the sequence does not take place in an appreciable amount until the preceding member has first been produced in considerable concentration, and partly because Coulomb effects become increasingly inhibitory as larger and larger values of Z are reached, even at temperatures somewhat above 10^9 degrees.

The nuclei produced in the α process stand out in abundance above other neighboring nuclei. This is made clear in Fig. VI,1 which is discussed in Sec. VI in connection with the neutron-capture *s* process which synthesizes these nuclei. Mg^{24}, Si^{28}, etc., are much too abundant to have been produced in the *s* process. The α process is of course, very similar to helium burning. We differentiate the two processes on the basis that the alpha-particle sources are quite different in the two cases. Detailed calculations on the rates of the α-process reactions have been given by Hoyle (Ho54), by Nakagawa *et al.* (Na56), and by Hayakawa *et al.* (Ha56b). We estimate that a progression in

TABLE III,4. Relative abundances of nuclei produced in the α process (Su56).

	$(Ne^{20})^a$	Mg^{24}	Si^{28}	S^{32}	A^{36}	Ca^{40}	Ca^{44}	Ti^{48}
Q_α (Mev)	(9.31)	10.00	6.94	6.66	7.04	5.28	9.40	9.32
Abundance	(8.4)	0.78	1.00	0.39	0.14	0.052	0.0011	0.0019

a Produced in helium burning; included for comparison.

temperature from 10^9 to 3×10^9 degrees, either in various regions or at various times in a star, will suffice to bring the rates of the (α,γ) reactions from Mg^{24} to Ca^{40} into reasonable correspondence.

The α process is, of course, an oversimplification as temperatures near 3×10^9 degrees are considered. Interactions between the heavier nuclei themselves must also be taken into account. It is interesting to note that experimental information on the cross sections of these interactions are already becoming available from research with high-energy accelerators equipped with ion sources capable of producing heavy ions (Ha56a).

One interesting elaboration of the α process has not been previously discussed in the literature. The straightforward (α,γ) reactions involving stable nuclei terminate at Ca^{40} since Ti^{44} with an electron capture lifetime of approximately 1000 years (Hu57b) is produced next. We have

$$_{20}Ca^{40}+{}_2He^4\rightarrow{}_{22}Ti^{44}+\gamma$$

$$_{22}Ti^{44}+e^-\rightarrow{}_{21}Sc^{44}+\nu_+ \qquad \tau_{\frac{1}{2}}\sim1000\text{ years}$$

$$_{21}Sc^{44}\rightarrow{}_{20}Ca^{44}+\beta^++\nu_+ \qquad \tau_{\frac{1}{2}}=3.9\text{ hours.}$$

If the alpha-capture lifetimes are greater than 1000 years, then the foregoing decays occur before Ti^{44} captures another alpha particle. Then one has

$$_{20}Ca^{44}+{}_2He^4\rightarrow{}_{22}Ti^{48}+\gamma,$$

the Ti^{48} being stable. The important point is that Ti^{48} is produced in an interaction with approximately the same barrier height as in the case of Ca^{44} and thus these two should be of roughly equal abundance. This is just the case within a factor of two (favoring Ti^{48}!) in spite of the rapid decrease in abundance evident from Si^{28} to Ca^{44}. Some Ti^{48} may also be produced in the *e* process discussed in the next section.

A corollary of our argument is that the time-scale of the α process must exceed 1000 years at Ti^{44}. If Ti^{44} captured an alpha particle it would produce Cr^{48} which decays through V^{48} to Ti^{48}. The unconsumed Ti^{44} would eventually decay to Ca^{44}, and in this case one would expect Ca^{44} to be considerably more abundant than Ti^{48}.

E. Succession of Nuclear Fuels in an Evolving Star

Starting with primeval hydrogen condensed into stars, pure hydrogen burning, pure helium burning, and finally the α process successively take place at the stellar center and then move outward in reaction zones or shells. When the star first contracts the generation of energy by hydrogen burning develops internal pressures which oppose gravitational contraction, and the star is stabilized on the main sequence at the point appropriate to its mass. Similarly, the generation of energy in helium burning should lead to a period of

relative stability during the red-giant stage of evolution. It is assumed that mixing does not occur. This is substantiated by the fact that, as hydrogen becomes exhausted in the interior, the star evolves off the main sequence which is the location of stars with homogenous interiors.

At the end of helium burning most of the nuclear binding energy has been abstracted, and indeed the cycle of contraction, burning, contraction . . . must eventually end when the available energy is exhausted, that is, when the most stable nuclei at the minimum in the packing fraction curve are reached, near Fe^{56}. If a star which condensed originally out of pure hydrogen remains stable, it eventually forms the iron-group elements at its center, and this "iron core" continues to grow with time until gravitation, unopposed by further energy generation, leads eventually to a violent instability. This problem of the ultimate instability is discussed, but not solved, in Sec. XII. At this point it suffices only to emphasize that the instability may result in the ejection of at least part of the "iron core" and its thin surrounding shells of lighter elements into the interstellar medium and that in this way a reasonable picture of the production of the abundance peak at the iron-group elements can be formulated. Production of the iron group of elements requires temperatures near 4×10^9 degrees, at which statistical equilibrium is reached, as outlined in Sec. II and discussed in detail in Sec. IV (the *e* process).

F. Burning of Hydrogen and Helium with Mixtures of Other Elements; Stellar Neutron Sources

In the previous discussion we considered the effect of heating hydrogen and its reaction products to very high temperatures. First, the hydrogen is converted to helium, and the resulting helium is converted to C^{12}, O^{16}, and Ne^{20}. Then α particles released by (γ,α) reactions on the Ne^{20} build the α-particle nuclei Mg^{24}, Si^{28}, S^{32}, A^{36}, Ca^{40}, and also Ca^{44} and Ti^{48}. Finally, at very high temperatures, the latter nuclei are converted into the iron group. Further heating of the iron group, although of astrophysical importance, does not lead directly to any further synthesis. Thus all the remaining elements and isotopes must be provided for otherwise than by a cooking of pure hydrogen.

Very much more complicated reactions arise when we consider the cooking of hydrogen and helium mixed with small concentrations of the elements already provided for, e.g., C^{12}, Ne^{20}, Fe^{56}. It is easy to see how such admixtures can arise. Since stars eject the products of nuclear synthesis into the interstellar gas it seems highly probable that only the "first" stars can have consisted of pure hydrogen. The results of hydrogen cooking in such stars would follow the lines described above. But once the interstellar gas was contaminated by this first cooking, nuclear processes would operate on hydrogen which contains impurities. The eventual hydrogen exhaustion will lead to helium burning with impurities. As we shall see later, the presence of other nuclei can lead to highly important effects.

In addition, hydrogen and helium may in some cases become adulterated with impurities even in the "first" stars. For example C^{12}, Ne^{20} built in the inner central regions of such stars may be circulated into the outer hydrogen envelopes. The question whether such mixing occurs, and astrophysical observations bearing on the problem, is considered in Secs. XI and XII.

(*1*) *CN cycle*

When C^{12} produced in helium burning is mixed with hydrogen at high enough temperatures, hydrogen is converted to He^4 by the CN cycle in addition to the *pp* chain previously considered. The implications for energy generation in hot main-sequence stars have been considered by numerous authors since Bethe (Be39) and von Weizsäcker (We38). The reactions of the CN cycle are

$$C^{12}(p,\gamma)N^{13}(\beta^+\nu_+)C^{13}$$
$$C^{13}(p,\gamma)N^{14}$$
$$N^{14}(p,\gamma)O^{15}(\beta^+\nu_+)N^{15}$$
$$N^{15}(p,\alpha)C^{12}.$$

These four reactions produce C^{13}, the heavier stable isotope of carbon, and the two stable forms of nitrogen. For C^{12}, C^{13}, and N^{14} the (p,α) reaction is not exothermic and only the (p,γ) reaction occurs. At N^{15}, the (p,α) reaction becomes exothermic and much more rapid than the (p,γ) reaction, which serves only as a small leak of material to O^{16}. The $N^{15}(p,\alpha)$ reaction reproduces the original C^{12} and a true cycle of reactions is established. Not only does this give rise to the catalytic conversion of hydrogen into helium until hydrogen is exhausted, but it also results in the carbon and nitrogen isotopes *not* being consumed in hydrogen burning.

The cross sections of the CN-cycle reactions have been under experimental investigation in the Kellogg Radiation Laboratory for some years and a review of the reaction rates in stars as known up to 1954 is included in the earlier paper by Fowler (Fo54), with numerical computations by Bosman-Crespin *et al.* (Bo54). New measurements of the CN-cycle reactions in the 100-kev range of interaction energies are now underway by William A. S. Lamb and Ross E. Hester at the Livermore Radiation Laboratory, who are using an ion source capable of delivering several hundred milliamperes of protons in this energy range, a factor of almost 1000 over the currents used in previous experiments. Results on $C^{12}(p,\gamma)$ are already published (La57) and results have also been obtained on $N^{14}(p,\gamma)$ (La57a). These are the two slowest reactions in the CN cycle and are the ones primarily involved in the over-all rate of this reaction.

In Figs. III,4 and III,5 are shown the yield and

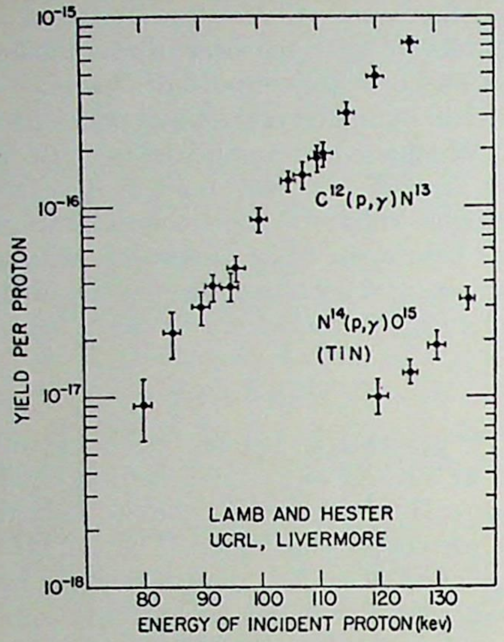


FIG. III,4. Yields of the $C^{12}(p,\gamma)$ and $N^{14}(p,\gamma)$ reactions as a function of bombarding energy obtained by Lamb and Hester (La57) at Livermore. The $N^{14}(p,\gamma)$ results were obtained using a titanium nitride target. In results not illustrated here the yield has been measured down to 100 kev where it has the very low value of only one reaction per 10^{18} incident protons.

cross-section curves for $C^{12}(p,\gamma)$ obtained by Lamb and Hester. For comparison, points obtained originally by Hall and Fowler (Ha50) and by Bailey and Stratton (Ba50) are shown. All of the observations are seen to fit a modification due to Thomas (Th52) of the Breit-Wigner dispersion formula using parameters of the resonance at 456 kev. The new results confirm the original measurements, but more importantly they extend the measurements to cross sections lower by a factor of ten and they do this with increased precision, so that the extrapolation to the lower temperatures relevant to stellar interiors ($E_o \sim 25$ kev) can be made with much more confidence. Their cross section $\sigma = 5.2 \times 10^{-11}$ barn at 90 kev is a representative one and yields

$$S = \left(\frac{12}{13} \times 90 \times 5.2 \times 10^{-11}\right) \exp\{+188.4/(90)^{\frac{1}{2}}\}$$

$$= 1.7 \text{ kev barns.}$$

Extrapolation to stellar energies yields

$$S_o = 1.2 \pm 0.2 \text{ kev barns for } C^{12}(p,\gamma)N^{13}$$

as found by Hall and Fowler (Ha50), so that no change in the numerical calculations of Bosman-Crespin *et al.* is required.

Experimental results for the reaction $N^{14}(p,\gamma)$ have also been obtained and are also shown in Figs. III,4 and III,5. Lamb and Hester find a representative value $\sigma = 7.0 \pm 1.0 \times 10^{-11}$ barn at 125 kev. This corresponds to a cross-section factor

$$S = \left(\frac{14}{15} \times 125 \times 7.0 \times 10^{-11}\right) \exp\{+219.8/(125)^{\frac{1}{2}}\}$$

$$= 2.8 \text{ kev barns.}$$

Their value, $\sigma = 0.9 \times 10^{-11}$ barn, at 100 kev is the smallest charged particle cross section measured up to the present time.

This and their other values correspond to a cross-section factor at stellar energies equal to $S_o = 3.2$ kev barns. In the Kellogg Radiation Laboratory Pixley (Pi57) has made measurements on this reaction from 220 kev up to 650 kev. His results can be interpreted as indicating that a narrow resonance at 278 kev, with width $\Gamma = 1.7$ kev, is superimposed on a non-resonant background which increases with the energy just as the calculated Coulomb barrier penetration factor times the usual $1/v$-term. The nonresonant cross section varies from 1.44×10^{-7} barn at 450 kev to 4.74×10^{-7} barn at 650 kev. The measured cross sections correspond to a cross-section factor at low energies given by $S_o = 2.8$ kev barns. The two sets of experiments indicate that the nonresonant background is characterized by a relatively constant S from 130 to 650 kev. Averaging the two values given above we have, with a reasonable allowance for systematic errors,

$$S_o = 3.0 \pm 0.6 \text{ kev barns for } N^{14}(p,\gamma)O^{15}.$$

This result is 12% of the value derived from the measurements of Woodbury, Hall, and Fowler (Wo49) near 124-kev bombarding energy, and confirms the finding of Duncan and Perry (Du51) that this old value could not be attributed to the effects of known resonances from 278 kev to 2.6 Mev. Fowler (Fo54) attributed the large value at 124 kev to the resonance effects of a bound level at 6.84 Mev in O^{15}, which is 543 kev below the $N^{14}+p$ threshold, and Bosman-Crespin *et al.* (Bo54) accordingly used, near $T = 10^7$ degrees, the value $S_o = 33$ kev barns for $N^{14}(p,\gamma)$. There is now no basis for this value. Accordingly, the tabulated values of Bosman-Crespin *et al.* for $N^{14}(p,\gamma)O^{15}$ must be multiplied by ~ 0.09. On this basis the $N^{14}(p,\gamma)$ reaction is very much slower than $C^{12}(p,\gamma)$.

Lamb and Hester (La57) obtain preliminary values for the cross section for $C^{13}(p,\gamma)$ which are about twice those which were obtained by Woodbury and Fowler (Wo52). Until this discrepancy is cleared up it is recommended that the reaction rates given by Fowler (Fo54) be employed with some caution. No new measurements are available on $N^{15}(p,\alpha)$ and the rates given by Fowler (Fo54) can be employed. Cross-section factors given by Fowler (Fo54) are:

$$S_o = 6.1 \pm 2 \text{ kev barns for } C^{13}(p,\gamma)N^{14},$$

$$S_o = 1.1 \pm 0.3 \times 10^5 \text{ kev barns for } N^{15}(p,\alpha)C^{12}.$$

In spite of the concordance of the new off-resonant measurements for $N^{14}(p,\gamma)$ there still exists the vexing problem as to whether resonance occurs in $N^{14}(p,\gamma)O^{15}$ in the region of stellar energies corresponding to an excitation energy in O^{15} near 7.37 ± 0.03 Mev. Known levels in O^{15} near this excitation occur at 6.84, 6.90, 7.61 ($\frac{1}{2}+$), and 8.33 Mev ($\frac{3}{2}+$). The spin and parity assignments for the last two levels come from studies of the $N^{14}(p,p)$ scattering by Pixley (Pi57) and by Hagedorn *et al.* (Ha57). The 7.61-Mev level is the level investigated in detail at 278-kev bombarding energy by Pixley. It is too narrow and too far removed to influence the cross section at 7.37 kev. In the mirror nucleus N^{15}, which has been investigated more completely, levels occur at 7.16 ($\frac{7}{2}+$), 7.31 ($\frac{3}{2}+$), 7.58 ($\frac{5}{2}+$?), 8.32 ($\frac{1}{2}+$), and 8.57 Mev ($\frac{3}{2}+$). The spin and parity assignments are the results of a recent comparison of shell model calculations with experiment by Halbert and French (Ha57a). In the light of the spins and parities and of the expected large level shifts (Th52) it can be convincingly argued that the last two levels correspond to the last two indicated above for O^{15}. One of the other three has not been found in O^{15}. Considerable work has been done on this problem in the last few years, but no new level has been found in this region in O^{15}, and a possible explanation lies in the level shift phenomenon. However, the research has not been exhaustive and it is still possible that a level falls in the range 7.37 ± 0.03 Mev, where it would serve as a resonance at stellar energies and thus would possibly enhance by a considerable amount the $N^{14}(p,\gamma)$ reaction rate in main-sequence stars.

If resonance in $N^{14}(p,\gamma)$ occurs then the reaction could be much more rapid than $C^{12}(p,\gamma)$, in which case the over-all rate of the CN cycle is determined by this last process. In this case the energy generation is that given as ϵ_C by Bosman-Crespin *et al.* (Bo54),

$$\epsilon_C=(3.18\pm0.54)\alpha_C x_H x_C\times10^{27}\ \text{ergs g}^{-1}\ \text{sec}^{-1},$$

where x_C is the fraction by mass of C^{12}. This is about 81% of the total carbon, C^{12} plus C^{13}, and there will be very little nitrogen. Thus $x_C\approx0.8x_{CN}$. The quantity α_C is given by

$$\alpha_C=\rho f_C(T_6^{-\frac{2}{3}}+0.017)\exp(-136.9T_6^{-\frac{1}{3}})$$
$$\simeq1.4\times10^{-25}\rho f_C\left(\frac{T_6}{15}\right)^{17.9}\quad\text{near } T_6=15,$$

with the electron shielding factor given by Salpeter's strong screening formula (Sa54) as $f_C=\exp(0.98\rho^{\frac{1}{2}}/T_6)$.

The mean lifetime of hydrogen becomes

$$\tau_C(\text{H})=1.90(x_H/\epsilon_C)\times10^{11}\ \text{years}$$
$$=0.60(\alpha_C x_C)^{-1}\times10^{-16}\ \text{year},$$

while for C^{12} the mean *cycle* time is

$$\tau_H(C^{12})=0.63(x_C/\epsilon_C)\times10^{11}\ \text{years}$$
$$=0.20(\alpha_C x_H)^{-1}\times10^{-16}\ \text{year}.$$

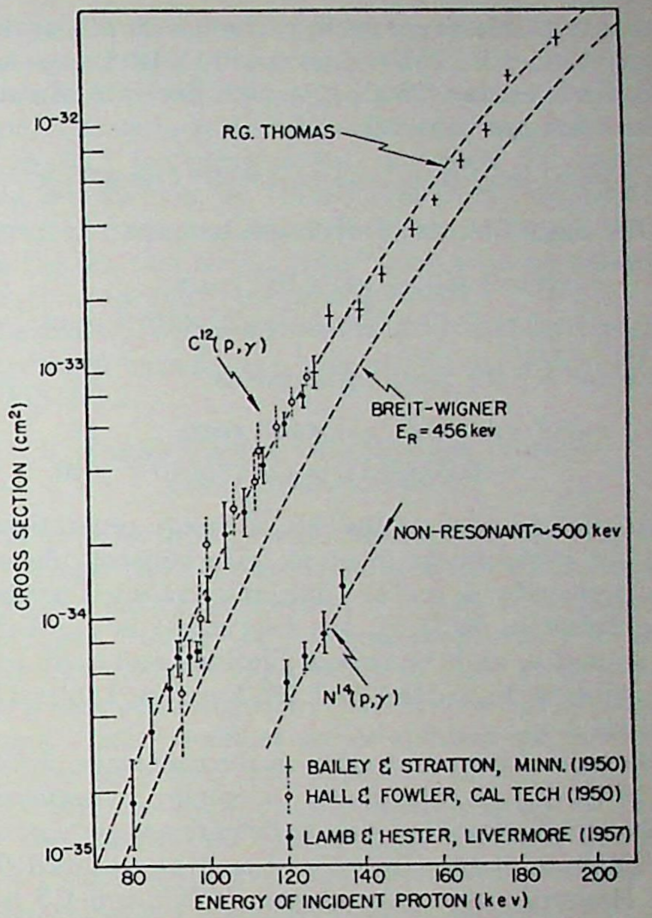


FIG. III,5. The cross sections of the $C^{12}(p,\gamma)$ and $N^{14}(p,\gamma)$ reactions at low bombarding energies. The early results on $C^{12}(p,\gamma)$ of Bailey and Stratton (Ba50) at the University of Minnesota, and of Hall and Fowler (Ha50) at the California Institute of Technology are compared with the new and more accurate results of Lamb and Hester (La57) obtained at the University of California Radiation Laboratory, Livermore. A modification of the Breit-Wigner dispersion formula by Thomas (Th52) based on the parameters of the resonance at 456 kev is found to fit the experimental result very well. Note the extremely low cross section of 10^{-11} barn measured experimentally by Lamb and Hester.

On the other hand, if resonance does not occur in $N^{14}(p,\gamma)$ then we must use the new value for S_o already given. Ignoring all small correction terms and defining

$$\alpha_N=\rho f_N T_6^{-\frac{2}{3}}\exp(-152.3T_6^{-\frac{1}{3}})$$
$$\simeq2.5\times10^{-28}\rho f_N\left(\frac{T_6}{15}\right)^{19.9}\quad\text{near } T_6=15,$$

with the electron shielding factor given by Salpeter's strong screening expression as $f_N=\exp(1.11\rho^{\frac{1}{2}}/T_6)$, then the reaction rate of the CN cycle per gram of material per second becomes

$$r_N=(1.96\pm0.39)(\alpha_N x_H x_N)\times10^{32}\ \text{reactions g}^{-1}\ \text{sec}^{-1},$$

where x_N is the fraction by mass of N^{14}. In this case almost all of the carbon and nitrogen will be in the form of N^{14}. Thus $x_N\approx x_{CN}$. The mean reaction rate per proton with four protons consumed per cycle is

$$p_N(\text{H})=-\frac{1}{x_H}\frac{dx_H}{dt}=(1.31\pm0.26)(\alpha_N x_N)\times10^{9}\ \text{sec}^{-1}.$$

The energy release per cycle, excluding the 6% neutrino energy loss, is 25.04 Mev $=4.011\times10^{-5}$ erg using Wapstra's masses (Wa55), so that the rate of energy generation can be written

$$\epsilon_N=(0.786\pm0.16)(\alpha_N x_H x_N)\times10^{28}\text{ ergs g}^{-1}\text{ sec}^{-1}.$$

The mean lifetime of hydrogen becomes

$$\tau_N(\text{H})=1.90(x_H/\epsilon_N)\times10^{11}\text{ years}$$
$$=(2.42\pm0.48)(\alpha_N x_N)^{-1}\times10^{-17}\text{ year},$$

while for N^{14} the mean *cycle* time is

$$\tau_H(\text{N})=0.55(x_N/\epsilon_N)\times10^{11}\text{ years}$$
$$=(0.70\pm0.14)(\alpha_N x_H)^{-1}\times10^{-17}\text{ year}.$$

At the present time the rate of energy generation by the CN cycle can be taken as lying between the wide limits given by ϵ_N and ϵ_C. Since resonance has not been established in $N^{14}(p,\gamma)$, the best guess is ϵ_N. Values for ϵ_C and ϵ_N and the various lifetimes and cycle times are given in Figs. III,1 and III,2 and in Tables III,1 and III,2.

Because of the uncertainty in the resonance problem in $N^{14}(p,\gamma)$ it is reasonable to employ astrophysical arguments in search of a more definite answer concerning the rate of the CN cycle. Let it be assumed that the observed cosmic abundances of N^{14} and C^{13} have been produced at equilibrium at temperature T from C^{12} in the CN cycle. Then if $S_o=33$ kev barns (Fo54) for $N^{14}(p,\gamma)$ and $S_o=6.1$ kev barns (Fo54) for $C^{13}(p,\gamma)$ it can be calculated that the relative numbers of nuclei are given by

$$N^{14}/C^{13}=[133/T_6]^2.$$

The relative abundance given by Suess and Urey (Su56) is

$$N^{14}/C^{13}=168.$$

The temperature giving this ratio is approximately 10^7 degrees. At this temperature the C^{12} cycle time is 3×10^9 years for $\rho x_H=100$ g/cm³, so conversion of C^{12} to C^{13} or N^{14} will occur only in old stars.

With the lower limit of $S_o=3.0$ kev barns for N^{14} (p,γ) it is found that

$$N^{14}/C^{13}=[440/T_6]^2.$$

The temperature corresponding to the observed relative abundance is $\sim T=3.4\times10^7$ degrees. This value is close to the temperature believed to hold in the hydrogen-burning shells surrounding the helium cores of red giant stars (Ho55). It is perhaps reasonable astrophysically to argue that this constitutes the last situation in which C^{13} and N^{14} are in equilibrium under hydrogen-burning. The C^{13} and N^{14} from these shells will be ejected into space when instability arises in the stellar interior. C^{13} and N^{14} which remain in the helium region after the exhaustion of hydrogen are consumed in helium burning by $C^{13}(\alpha,n)$ and $N^{14}(\alpha,\gamma)$. On the basis, then, of this argument, the lower limit for S_o would seem to be the preferred value.

With the new value for S_o for $N^{14}(p,\gamma)$ it is found that the CN cycle rate will exceed that for the *pp* chain only in stars with central temperatures over 18×10^6 degrees. The existence of nitrogen-rich peculiar stars, as discussed in Sec. XI, is to be expected on this basis since at equilibrium in the CN cycle we now have

$$N^{14}/C^{12}=[200/T_6]^2$$

so that $N^{14}\gg C^{12}$ at all reasonable temperatures ($T_6<40$) for hydrogen burning. Stars in which carbon from the helium-burning cores has been passed slowly enough through hydrogen-burning zones and then mixed into the outer atmosphere will show a much greater over-abundance of nitrogen than of carbon. The C^{12}/C^{13} ratio by number should be 4.6 for the carbon in such stars. On the other hand, stars in which C^{12}, O^{16}, and Ne^{20} from helium-burning zones were mixed quickly without hydrogen-burning into the envelope would show anomalous abundances of these isotopes and not necessarily of N^{14} or C^{13}. The ratio of C^{12}/C^{13} might be quite large in such stars. These points have been emphasized by Greenstein (Gr54a); see also Sec. XI A.

Finally, with the new nonresonant value for N^{14}-(p,γ) the CN cycle is very slow in the sun. The recent inhomogeneous model calculations of Schwarzchild, Howard, and Härm (Sc57) which use only the *pp* chain need not be modified to include CN cycle processing of hydrogen except for the cases involving rather low values of the initial hydrogen content. D. E. Osterbrock (private communication) has kindly made differential corrections to the calculations of Schwarzschild *et al.* using the *pp* reaction and $N^{14}(p,\gamma)$ reaction rates given in this paper. He finds $T_c=15.4\times10^6$ degrees and $\rho_c=147$ g/cm³ for an original hydrogen content of 80% by mass. These values replace $T_c=14.8\times10^6$ degrees and $\rho_c=132$ g/cm³. For the present values $x_H=0.30$ and $x_{CN}=0.003$ at the solar center he finds $\epsilon_{pp}=5.25$ ergs g^{-1} sec^{-1} and $\epsilon_{CN}=0.20$ erg g^{-1} sec^{-1}. Thus the CN-cycle energy generation is only about 4% of that of the *pp* chain at the center and is less than this value on the average.

(2) *Other Hydrogen-Burning Reactions*

The reactions of protons and other light nuclei have been discussed by Fowler (Fo54) and by Salpeter (Sa55) and most recently by Marion and Fowler (Ma57). The rapid destruction of the lithium, beryllium, and boron isotopes in hydrogen burning, the end result being the production of helium, is well known and will not be discussed here, though some of the problems which this raises for synthesis of these elements are discussed in Sec. X.

At the temperatures occurring in the hydrogen-burning shells of red giant stars O^{16} is converted to

N^{14} by

$$O^{16}(p,\gamma)F^{17}(\beta^+\nu_+)O^{17}$$
$$O^{17}(p,\alpha)N^{14}+1.20 \text{ Mev (exothermic).}$$

The N^{14} is then processed by the CN cycle. When the small leakage due to $N^{15}(p,\gamma)O^{16}$ from the CN cycle is taken into account with the above reactions one has a generalized CNO cycle.

For $O^{16}(p,\gamma)$ Tanner and Pixley (Ta57a) give the preliminary value

$$S_o=4\pm 2 \text{ kev barns for } O^{16}(p,\gamma)$$

so that the lifetime for O^{16} in hydrogen burning becomes

$$\tau_H(O^{16})=0.5(\alpha_O x_H)^{-1}\times 10^{-17} \text{ year,}$$

where

$$\alpha_O=\rho T_6^{-\frac{2}{3}}\exp(-166.7T_6^{-\frac{1}{3}}).$$

Here we have neglected the weak electron shielding expected to hold in high-temperature, low-density shells. Since $O^{16}(p,\gamma)$ is followed rapidly by $O^{17}(p,\alpha)$ and by hydrogen processing in the CN cycle, the hydrogen lifetime will be the same as that for O^{16}. Whenever O^{16} burning is not complete, a small amount of O^{17} $(\sim 10^{-3}\times O^{16})$ remains in equilibrium with O^{16}. This may be sufficient to account for the small concentration of $O^{17}=4\times 10^{-4}\times O^{16}$ observed in the oxygen isotope ratios. Apparently, not all of the O^{16} and O^{17} have been destroyed in the hydrogen-burning shells of the giant stars during the relatively short life of this evolutionary stage. On the other hand, $O^{18}=2\times 10^{-3}\times O^{16}$ is not produced in sufficient abundance by $O^{17}(p,\gamma)F^{18}(\beta^+\nu_+)O^{18}$ since the (p,α) reaction is much faster than the (p,γ) for O^{17}.

The burning of Ne^{20} in hydrogen is followed by the NeNa cycle which is somewhat similar to the CN cycle,

$$Ne^{2}\ (p,\gamma)Na^{21}(\beta^+\nu_+)Ne^{21}$$
$$Ne^{21}(p,\gamma)Na^{22}(\beta^+\nu_+)Ne^{22}$$
$$Ne^{22}(p,\gamma)Na^{23}$$
$$Na^{23}(p,\alpha)Ne^{20}.$$

New calculations based on recent experimental information have been made for the rates of the NeNa-cycle reactions by Marion and Fowler (Ma57). It is found that $Na^{23}(p,\alpha)$ is considerably faster than $Na^{23}(p,\gamma)$, so that little leakage from the cycle occurs to Mg^{24}. A method of obtaining the reduced nucleon emission widths for the corresponding levels of mirror nuclei is presented and is used in particular in the calculation of the $Ne^{20}(p,\gamma)$ reaction rate. It is found that a bound level in Na^{21} at $Ne^{20}+p-26$ kev markedly enhances the rate of this reaction and that $S_o\approx 12$ kev barns at the temperatures of hydrogen-burning shells. Results of preliminary measurements by Pixley, Hester, and Lamb (Pi57a) increase this to

$$S_o\simeq 75 \text{ kev barns for } Ne^{20}(p,\gamma)Na^{21}.$$

Substantial quantities of Ne^{21} are produced at the elevated temperatures occurring just before the hydrogen is exhausted in such shells. This will be true only if the carbon, nitrogen, and oxygen isotopes are relatively rare so that hydrogen is not processed too rapidly before the Ne^{20} can be burned. The Ne^{21} thus formed becomes important in the subsequent helium burning when the reaction $Ne^{21}(\alpha,n)Mg^{24}$ may serve as a source of neutrons for heavy element synthesis.

The mean lifetimes or cycle times of the light nuclei in hydrogen-burning are summarized in Table III,2 and Fig. III,2.

(3) *Helium Burning with Reaction Products; Stellar Neutron Sources*

The primary products of pure helium burning are C^{12}, O^{16}, and Ne^{20}. When these are subjected to further hydrogen burning, production of all of the stable isotopes from C^{12} to Na^{23}, with the exception of N^{15}, O^{18}, and F^{19}, is accounted for. Production of these three exceptions is treated at the end of this section. In hydrogen burning, cyclic processes are established which transform the hydrogen into helium without consuming the cycling, catalytic nuclei. Exceptions are O^{16} and O^{17}; these nuclei are consumed in hydrogen burning so it is clear that a proportion of these isotopes originally produced has not been subjected to terminal hydrogen burning.

Exhaustion of the hydrogen by the cyclic processes eventually means that the unconsumed isotopes of carbon, nitrogen, neon, sodium, and perhaps oxygen, will be mixed with helium. Upon contraction the mixture comes to a temperature at which helium begins to interact with these nuclei. Numerous reactions occur but by far the most significant are the exothermic (α,n) reactions which provide a source of neutrons for the *s* and *r* processes by which the heavy elements can be synthesized.

The first stellar neutron source was proposed by Greenstein (Gr54) and by Cameron (Ca54, Ca55), namely the *exothermic* reaction:

$$C^{13}(\alpha,n)O^{16}+2.20 \text{ Mev.}$$

To this, Fowler *et al.* (Fo55) added the *exothermic* (α,n) reactions involving the $A=4n+1$ nuclei, *viz.*,

$$O^{17}(\alpha,n)Ne^{20}+0.60 \text{ Mev}$$
$$Ne^{21}(\alpha,n)Mg^{24}+2.58 \text{ Mev}$$
$$Mg^{25}(\alpha,n)Si^{28}+2.67 \text{ Mev}$$

and the special case,

$$Mg^{26}(\alpha,n)Si^{29}+0.04 \text{ Mev.}$$

The next $A=4n+1$ reaction, $Si^{29}(\alpha,n)S^{32}$, is *endothermic* by 1.53 Mev, and $S^{33}(\alpha,n)A^{36}$ is *endothermic* by 2.0 Mev.

Insofar as neutron sources for the *s* process in red giants are concerned, we must confine our attention to C^{13} and Ne^{21}. Very little O^{17} remains for helium burning

because it is destroyed by $O^{17}(p,\alpha)$ in hydrogen burning. Mg^{25} and Mg^{26} are not produced in hydrogen burning because the hydrogen is consumed in other processes before $Mg^{24}(p,\gamma)$ becomes operative.

Marion and Fowler (Ma57) have recently discussed the rates of the $C^{13}(\alpha,n)$ and $Ne^{21}(\alpha,n)$ reactions. For the $C^{13}(\alpha,n)$ reaction they find $S_o=2.1\times10^{11}T_6^{-\frac{2}{3}}$ kev barn so that the C^{13} lifetime is

$$\tau_{\alpha n}(C^{13})=4.9\times10^{-28}\frac{T_6^{2}}{\rho x_\alpha}\exp(323T_6^{-\frac{1}{3}})\text{ years.}$$

For the $Ne^{21}(\alpha,n)$ reaction they find $S_o=1.6\times10^{13}$ kev barns so that the Ne^{21} lifetime is

$$\tau_{\alpha n}(Ne^{21})=5.6\times10^{-30}\frac{T_6^{\frac{2}{3}}}{\rho x_\alpha}\exp(468T_6^{-\frac{1}{3}})\text{ years.}$$

A reaction rate of importance in the following discussion is that of $N^{14}(\alpha,\gamma)$, for which it was found that $S_o=1.2\times10^{7}T_6^{-\frac{2}{3}}$ kev barn and

$$\tau_{\alpha\gamma}(N^{14})=8.2\times10^{-24}\frac{T_6^{2}}{\rho x_\alpha}\exp(360T_6^{-\frac{1}{3}})\text{ years.}$$

The above lifetimes are indicated graphically in Fig. III,3 and are tabulated in Table III,3.

Neither $C^{13}(\alpha,n)$ nor $Ne^{21}(\alpha,n)$ are wholly free from objection as stellar neutron sources. Cameron (Ca55) and Fowler *et al.* (Fo55) have discussed the problems which arise if the $C^{13}(\alpha,n)O^{16}$ reaction is taken as the principal source of neutrons for heavy element synthesis in helium burning during the giant stage of a star's evolution. The essential difficulty lies in the fact that only a small amount of C^{13} is produced at equilibrium in the carbon-nitrogen cycle; $C^{13}/C^{12}=1/4.6$ by number at equilibrium. As a result too few neutrons are produced when C^{13} begins to interact with the helium. The cosmic abundance ratio $C^{12}/Fe^{56}=6.4$ implies that $6.4/4.6=1.4$ neutrons will become available per iron nucleus and these will only be sufficient to build nuclei slightly heavier than Fe^{56}. Another difficulty is the fact that N^{14} is the most abundant of the isotopes at equilibrium in the CN cycle; $N^{14}/C^{13}=168$ even at the high temperature of 3.4×10^{7} degrees for a hydrogen-burning shell. The reaction $N^{14}(n,p)C^{14}$ consumes a large fraction of the neutrons produced in the $C^{13}(\alpha,n)O^{16}$ reaction. These difficulties can be avoided, as emphasized by Cameron, with the $C^{13}(\alpha,n)O^{16}$ reaction as the neutron source, if it is postulated that considerable mixing between core and envelope takes place during the giant stage. In this case hydrogen from the envelope interacts with C^{12} produced by $3He^4\rightarrow C^{12}$ in the core and constantly replenishes the C^{13} supply but at such a rate that little N^{14} is produced by the $C^{13}(p,\gamma)N^{14}$ reaction. In addition if C^{12}, produced in the hot center of a core, is mixed into the cooler outer regions of the core where C^{13} is burning, then the C^{12} captures the protons from $N^{14}(n,p)$ and replenishes the C^{13}, thus permitting the full quota of neutrons to be made available for heavy element synthesis just beyond Fe^{56}.

Fowler *et al.* proposed the $Ne^{21}(\alpha,n)Mg^{24}$ reaction as an alternative source of neutrons which would avoid the foregoing difficulties without calling upon core-envelope mixing with its attendant stability problems. In this proposal it is assumed that Ne^{20}, previously produced during helium burning at an earlier stage of galactic evolution, is converted into Ne^{21} in the hydrogen-burning shells at $30–50\times10^{6}$ degrees surrounding the helium-burning cores of red-giant stars. When the shell hydrogen is converted into what then becomes core helium, the Ne^{21} interacts with helium to produce neutrons. The neutrons are captured by iron-group nuclei, also previously synthesized, to produce heavy elements. The atomic abundance ratio $Ne^{20}/Fe^{56}=14$ is taken as indicating that ~14 neutrons will become available per iron nucleus, a sufficient number for considerable heavy element synthesis. This suggestion thus requires that:

(i) The reactions $Ne^{20}(p,\gamma)Na^{21}(\beta^+\nu_+)Ne^{21}$ produce Ne^{21} faster than it is destroyed by the $Ne^{21}(p,\gamma)Na^{22}$ reaction.

(ii) The Ne^{21} is produced before the hydrogen mixed with the Ne^{20} is exhausted by the *pp* chain or the CN-cycle reactions.

(iii) The N^{14} is depleted by the $N^{14}(\alpha,\gamma)F^{18}$ reaction before the $Ne^{21}(\alpha,n)Mg^{24}$ reaction supplies the neutrons.

(iv) The Ne^{21} is consumed by $Ne^{21}(\alpha,n)Mg^{24}$ before helium is depleted in the core.

Because of the uncertainty in the reaction-rate calculations presented here, it is not possible to give unequivocal answers as to whether these conditions are satisfied; however, the following qualified statements can be made. With regard to (i) the results incorporated in Table III,2 and Fig. III,2 indicate that Ne^{21} is indeed produced from Ne^{20} faster than it is destroyed. If the NeNa cycle reaches equilibrium, then about 90% of the Ne^{20} could have been converted into Ne^{21} along with small amounts of Ne^{22} and Na^{23}. With regard to (ii), it is clear that the conversion of Ne^{20} into Ne^{21} will not occur before the hydrogen is exhausted by the CN cycle or oxygen reactions unless there is only a small concentration of carbon, nitrogen, and oxygen, perhaps 0.1% that of hydrogen by weight. This is $\sim8\%$ the normal abundance but could occur in some stars. With regard to (iii) it may be seen from Table III,3 and Fig. III,3 that N^{14} will be fairly well scoured out at low temperatures before Ne^{21} begins to interact, although there is again considerable uncertainty in this conclusion.

With regard to (iv), Table III,3 and Fig. III,3 show that $Ne^{21}(\alpha,n)$ will compete with other helium reactions only at elevated temperatures near 2×10^{8}

degrees in the core. Even then it is about a hundred times slower than the consumption of He^4 by $3He^4 \rightarrow C^{12}$ as long as helium is abundant. However, there is a factor of x_α^2 in the rate of $3He^4 \rightarrow C^{12}$ relative to $Ne^{21}(\alpha,n)$, where x_α is the concentration by weight of the helium. Thus when $x_\alpha < 0.1$, the Ne^{21} will be consumed. Indeed, we can expect that as x_α decreases the core temperature will rise to maintain the energy generation and eventually the last He^4 nuclei will rapidly scour out the Ne^{21} at high temperature. In any case, even if the Ne^{21} remains impervious to helium burning it will lose its loosely-bound neutron in the subsequent interactions of the medium-weight nuclei.

With the above considerations in mind a tentative picture of the generation of neutrons in red giant cores is as follows. We assume that the core will spend some 10^7 to 10^8 years at low temperature $(20-40\times 10^6$ degrees) while the star slowly progresses into the giant stage, some 10^5 years at 50 to 100×10^6 degrees, and then only a brief period, say 10^3 years, at 200×10^6 degrees as it rapidly exhausts helium and burns the middle weight elements. The C^{13} lifetime to the (α,n) reaction is about 10^5 years at 75×10^6 degrees, so that C^{13} will be consumed at this stage. Thus a few neutrons per Fe^{56} become available on a time-scale of about 10^5 years. We have seen in Sec. II that a time-scale $\lesssim 10^5$ years per neutron capture is demanded to produce the early part of the abundance distribution built by the *s* process; this is discussed further in Sec. VI. With the consumption of the C^{13} in its central regions, the core now heats to greater temperatures and rapidly processes He^4 by $3He^4 \rightarrow C^{12}$, etc. Toward the end of the helium burning, instability may well set in and mixing may occur throughout the core and with the envelope hydrogen. New C^{13} is produced, Ne^{21} burns, and a great flux of neutrons is produced in the final period of about 10^3 years. In Sec.II we showed that a shorter time-scale was necessary to account for the *s*-process abundance distribution for elements with $A \gtrsim 100$, where the atomic abundance curve becomes relatively flat. This is discussed further in Sec. VI, and in Sec. XI in connection with the observed overabundances of heavy elements in stars believed to have recently undergone *s*-process synthesis in their interiors.

Finally, we turn to neutron sources on a very short time-scale, $\sim$10 seconds, as in supernovae. A mechanism by which neutrons could arise in the hot envelope of a supernova following core collapse was proposed by Burbidge, Hoyle, Burbidge, Christy, and Fowler (Bu56). This mechanism is thought to occur in four stages:

(i) A collapse of the inner regions of a highly evolved star occurs leading to the conversion of gravitational energy into other forms of energy, particularly in the envelope of the star where it is converted into thermal energy. A possible cause of such a collapse is discussed in detail in Sec. XII. The imploding envelope reaches 2×10^8 degrees as a result of the conversion of gravitational energy into thermal energy. The density rises to $\sim 10^5$ g/cm^3.

(ii) The onset of reactions of the type $C^{12}(p,\gamma)N^{13}$, $O^{16}(p,\gamma)F^{17}$, $Ne^{20}(p,\gamma)Na^{21}$, $Mg^{24}(p,\gamma)Al^{25}$ take place in the outer parts of the star. These occur very rapidly at 2×10^8 degrees. The mean energy production from these reactions may be taken as approximately 2 Mev per proton. With a composition characteristic of a highly evolved star, consisting of approximately equal abundances (by number) of hydrogen, helium, and light nuclei, the energy released in these reactions is sufficient to raise the temperature of the material to a value $\sim 10^9$ degrees as the hydrogen is consumed. If all of the nuclear energy were converted into particle energy, a temperature of 6×10^9 degrees would be attained. This maximum temperature will not be reached, however, because of the conversion of energy into radiation, turbulence, growing magnetic fields and the general expansion of the eventually exploding envelope. However, 10^9 degrees will probably be reached during the initial implosion of the envelope into the region vacated by the collapsing core. A radiation temperature of $\sim 10^9$ degrees will result from the hydrogen consumption if $\rho = 10^5$ g/cm^3.

(iii) At this high temperature, alpha-particle reactions become important, and the crucial neutron-producing reaction is $Ne^{21}(\alpha,n)Mg^{24}$, which follows very rapidly on the 23-second beta decay of Na^{21}. The mean reaction time for Ne^{21} is of the order of 10^{-5} sec if $T \simeq 10^9$ degrees and $\rho \simeq 10^5$ g/cm^3. The (α,n) reactions on C^{13} and O^{17} are not operative under these conditions since N^{13} requires 14 min in the mean to decay to C^{13} and F^{17} requires 100 sec to decay to O^{17}. A series of neutron-producing reactions from combined hydrogen and helium burning on Ne^{20} and heavier nuclei are shown in Fig. III,6. All of the beta decays shown in this diagram are reasonably short. In general (p,γ) reactions compete with positron emission until proton addition is no longer possible and positron emission must occur.

(iv) The neutrons produced in hydrogen and helium burning are rapidly thermalized ($kT \simeq 100$ kev) and then they are captured primarily by the abundant nuclei of the iron group which were initially present in the envelope. We have supposed that their abundances at this stage total about 1% by mass. If the envelope was very underabundant in these elements the neutrons would be captured by the light elements. This rapid neutron capture process, the *r* process, will be discussed in detail in Secs. VII and VIII. For this discussion we require to know the neutron density. This will now be estimated.

The neutrons become available in the mean lifetime of Na^{21} (33 sec), or Mg^{22} (7 sec) and Al^{26} (10 sec). Thus a reasonable mean time for the neutron production during the imploding high-density stage is about 10

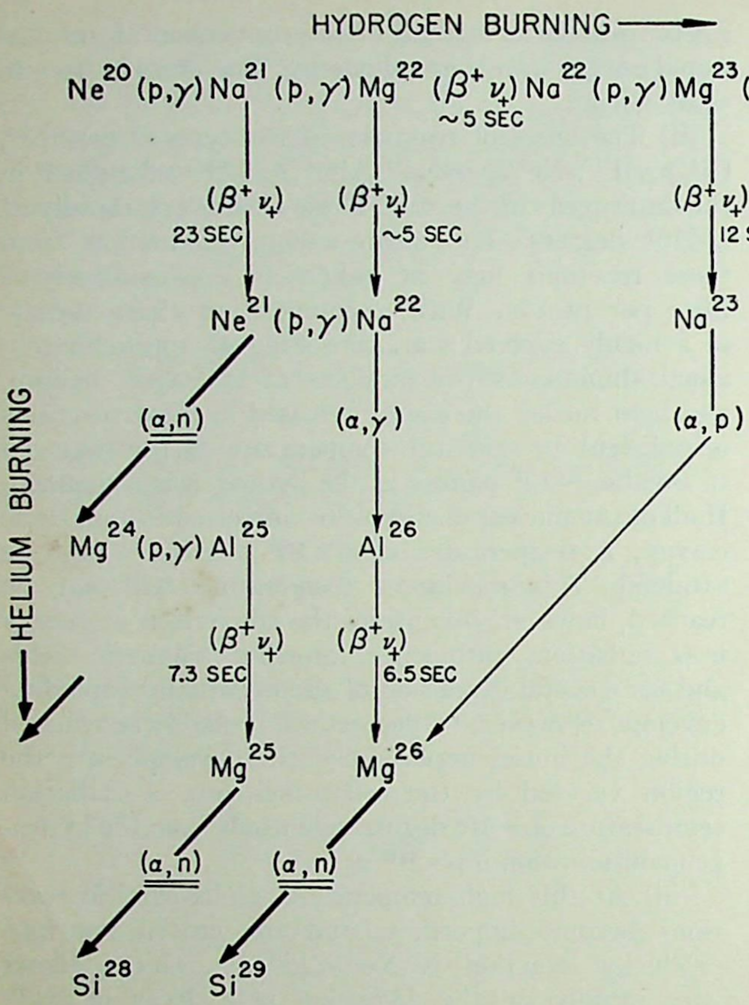


FIG. III,6. Neutron-producing reactions as a result of hydrogen and helium reactions in supernova envelopes. Note that all beta-decay lifetimes are quite short, and that neutrons are produced as long as hydrogen is 1 to 5 times as abundant as Ne^{20}.

sec. There will be one neutron produced for each Ne^{20} nucleus present, and on the assumption that there are approximately equal number densities of hydrogen, helium, and light nuclei, and that perhaps one-third of the light nuclei are together C^{12}, O^{16}, and Ne^{20}, each having equal abundance, the number density of Ne^{20} is $\sim 10^{22}\rho$. It has been earlier stated that a reasonable value for $\rho \simeq 10^5$ g/cm³, during the implosive stage. However, the neutron captures will occur during the later expanding stage when $\rho \simeq 100$ g/cm³, so that the neutron density during this stage will be $\sim 10^{24}$ neutrons/cm³. We assume that the radiation temperature remains at the high value of $\sim 10^9$ degrees during the capture process. As we shall see in Sec. VII, this means that captures occur until the neutron binding energy is reduced to ~ 2 Mev. Further captures must wait until beta emission occurs with lifetimes for the neutron-rich nuclei of about 3 sec. Each beta decay is followed by about 3 captures so that the total time for ~ 200 captures ($Fe^{56} \rightarrow Cf^{254}$) is ~ 200 sec. Thus the capture time is long as compared with the total neutron production time and we are justified in assuming that the initial neutron density is that given by the original Ne^{20} density.

To conclude, in our discussion of the r process in supernova envelopes we put the neutron density $n_n \simeq 10^{24}$/cm³ and the temperature $T \simeq 10^9$ degrees. The thermal neutron velocity at $kT \simeq 100$ kev is approximately 4×10^8 cm/sec, so that the neutron flux $(nv)_n \simeq 4 \times 10^{32}$ neutrons/cm² sec.

In Secs. X and XII brief mention is made of a further possible neutron source in a very highly collapsed stellar core.

(4) *Origin of N^{15}, O^{18}, and F^{19}*

The concentration of N^{15} given by the CN cycle is far too low, as compared with that of N^{14}, to explain the ratio $N^{15}/N^{14} = 1/270$ in the atomic abundance distribution. It is too low by a factor ~ 100, assuming the very slow rate for $N^{14}(p,\gamma)O^{15}$, discussed above in Sec. III F (1). Hence it appears that N^{15} must be produced by reactions other than those of the normal CN cycle. Let us consider the reaction $C^{13}(\alpha,n)O^{16}$, occurring in the presence of N^{14}, derived previously from the CN cycle, and also in the presence of a strong excess of C^{12} derived from helium burning. This requires mixing of material in a helium-burning core. Protons liberated by $N^{14}(n,p)C^{14}$ are, in the main, converted back to neutrons by the reactions $C^{12}(p,\gamma)N^{13}(\beta^+\nu_+)C^{13}$ and $C^{13}(\alpha,n)O^{16}$, since the C^{12}, in excess over its CN cycle abundance, will capture most of the protons. Nor are neutrons lost by the reaction $C^{12}(n,\gamma)C^{13}$, again because of the occurrence of the reaction $C^{13}(\alpha,n)O^{16}$. Thus the neutrons tend to cycle,

either through $N^{14}(n,p)C^{14}$, or through $C^{12}(n,\gamma)C^{13}$, the latter reaction being important because of the excess of C^{12}.

Neutrons are eventually lost, however, through capture in reactions with smaller probability. Important among these is $N^{14}(n,\gamma)N^{15}$. If the C^{12} excess is great enough for all protons, released by $N^{14}(n,p)C^{14}$, to be absorbed in $C^{12}(p,\gamma)N^{13}$, the ultimate result is to produce an N^{15} concentration comparable with that of the original C^{13} concentration. Since $C^{13}/N^{14}=1/168$, this implies the possibility of $N^{15}/N^{14}\sim 1/168$. As noted above, the isotopic ratio is $N^{15}/N^{14}=1/270$. It appears, therefore, that the present process, dependent on an excess of C^{12} becoming mixed with the products of the CN cycle in helium burning, is capable of providing for the origin of N^{15}.

At somewhat higher temperatures than are required to promote the above reactions, F^{19} is then produced by $N^{15}(\alpha,\gamma)F^{19}$. At the same time O^{18} is produced by $N^{14}(\alpha,\gamma)F^{18}(\beta^+\nu_+)O^{18}$. It would then follow that $F^{19}/O^{18}\sim N^{15}/N^{14}$. This is the case within a factor of 10.

To conclude, detailed mechanisms have been advanced for the production of the nuclei C^{12}, C^{13}, N^{14}, N^{15}, O^{16}, O^{17}, O^{18}, Fe^{19}, Ne^{20}, Ne^{21}, Ne^{22}, Na^{23} in combination hydrogen-helium-burning in stars. These mechanisms are illustrated graphically in Fig. I,2.

IV. *e* PROCESS

At temperatures above $\sim 3\times 10^9$ degrees all manner of nuclear processes occur in great profusion, i.e., (γ,α), (γ,p), (γ,n), (α,γ), (p,γ), (n,γ), (p,n) reactions as well as others involving heavier nuclei. Such reactions allow the conversion of a nucleus A, Z to a nucleus A', Z', even in cases in which Z and Z' are large; i.e., even though considerable Coulomb barriers may be involved in the conversion. This is the physical condition for statistical balancing to occur among the nuclei. In previous papers (Ho46, Ho54), it was shown that the abundances of the elements in the iron peak could be synthesized under conditions of temperature and density such that statistical equilibrium between the nuclei and the free protons and neutrons was achieved. In this section this question is re-examined using the more accurate nuclear and abundance data now available, and taking into account the appropriate excited nuclear states, the energies and spins for which are available now from the work of Way, King, McGinnis, and Lieshout (Wa55a).

Under conditions of statistical equilibrium the number density $n(A,Z)$ of the nucleus A, Z is given by

$$n(A,Z)=\omega(A,Z)\left[\frac{AMkT}{2\pi\hbar^2}\right]^{\frac{3}{2}}\frac{(n_n)^{A-Z}(n_p)^Z}{2^A}\times\left[\frac{2\pi\hbar^2}{MkT}\right]^{3A/2}\exp\frac{Q(A,Z)}{kT},\quad (1)$$

where n_n, n_p are the number densities of free neutrons and protons, M is the atomic mass unit, and $\omega(A,Z)$ is the statistical weight factor given by

$$\omega(A,Z)=\sum_r(2I_r+1)\exp(-E_r/kT),$$

where E_r is the energy of the excited state measured above the ground level, and I_r is the spin. $Q(A,Z)$ is the binding of the ground level of the nucleus A, Z, and is given by

$$Q(A,Z)=c^2[(A-Z)M_n+ZM_p-M(A,Z)],$$

where M_n, M_p, and $M(A,Z)$ are the masses of the free neutron, free proton, and nucleus A, Z, respectively. Writing $\theta=\log(n_p/n_n)$, and measuring T in units of 10^9 degrees (T_9), (1) can be expressed in the form

$$\log n(A,Z)=\log\omega(A,Z)+33.77+\tfrac{3}{2}\log(AT_9)+\frac{5.04}{T}Q(A,Z)+A(\log n_n-34.07-\tfrac{3}{2}\log T_9)+Z\theta.\quad (2)$$

This is the form given by Hoyle (Ho46) except that formerly the approximation was made of putting $\omega(A,Z)=2$, and a slight numerical change of 0.01 in the numerical term in the right-hand side of the equation has been made. Logarithms are to the base 10. The nuclear data required to calculate $Q(A,Z)$, $\omega(A,Z)$ are available, so that explicit values of $n(A,Z)$ for various A, Z can be worked out if n_n, θ, T_9 are specified. Thus it might seem, at first sight, as if abundances of the nuclei under statistical equilibrium are functions of three independent parameters. This is not the case, however. Two parameters are sufficient to determine the system. This can be understood best from taking ρ, the density, and T_9 as the independent parameters, although we shall see later that θ, T_9 give a more convenient choice of parameters from an arithmetical point of view. If ρ is given, the abundances calculated from the above equation must satisfy the condition

$$\sum M(A,Z)n(A,Z)+\sum M_pn_p+\sum M_nn_n=\rho,\quad (3)$$

the contribution of the electrons present in the system being neglected. The contribution of the free protons and free neutrons is also negligible for the values of ρ, T_9 used below, although values of ρ, T_9 could be chosen for which this would not be so. Equation (3) can be interpreted in a number of ways, for instance as determining $\log n_n$ in terms of ρ, T_9, θ. A second condition determining θ in terms ρ, T_9, arises in one or other of two alternative ways. If the time-scale is long enough, equilibrium between protons and neutrons is established through the operation of beta processes among the nuclei. This equilibrium is expressed by the equations

$$\left.\begin{aligned}&\frac{n_n}{n_p}=10^{-\theta}=\frac{n_e}{2}\left(\frac{2\pi\hbar^2}{m_ekT}\right)^{\frac{3}{2}}\exp(Q_o/kT),\\&n_e=\sum_{A,Z}Zn(A,Z)+n_p,\\&Q_o=c^2(M_p+m_e-M_n)=-0.78\text{ Mev},\end{aligned}\right\}\quad (4)$$

m_e being the electron mass, and n_e the electron number density.

If, on the other hand, the time-scale is not long enough to establish equilibrium between neutrons and protons, then (4) must be replaced by the condition that the ratio of the number of neutrons per unit volume (bound and free together), to the number of protons per unit volume must be a specified constant with a value determined by the initial state of the material. That is to say

$$\sum_{Z,A}(A-Z)n(A,Z)+n_n=(\text{specified constant})\times\Big[\sum_{Z,A}Zn(A,Z)+n_p\Big]. \quad (5)$$

Before proceeding to a numerical discussion of the above scheme of equations it is desirable to relate them very briefly to astrophysical considerations, particularly to the temperature evolution and time-scales shown schematically in Fig. II,4. So long as the time-scale is longer than a few months (4) are applicable, but in the final rise of temperature to the explosion point shown in Fig. II,4, and during the cooling that follows explosion, the time-scale is far too short to admit of the use of (4). During these phases (5) must be employed. An examination of (5) has shown that it requires θ to stay nearly constant at a value determined by the constant that appears on the right-hand side of the equation. This constant is in turn determined by the values of $n(A,Z)$, n_p, n_n that are operative at the last moment of applicability of (4).

We can now give a description of the physical and mathematical conditions that determine the abundances of the nuclei of the iron peak. Astrophysical reasons, discussed in Sec. XII, require a temperature evolution of the form shown in Fig. II,4 to take place in certain types of star. As previously pointed out, when T rises above about 3×10^9 degrees the statistical equations become applicable. So long as the time-scale is long enough, beta processes establish the equilibrium that satisfies (4). The decay of Mn^{56} with a half-life of 2.58 hours is important in this connection, as also is Co^{57} with a half-life of 267 days. The Co^{57} turns out to be a fairly abundant nucleus and its decay has the effect of changing protons to neutrons. The Mn^{56}, on the other hand, changes neutrons to protons. The time-scale for an appreciable interchange between neutrons and protons, or vice versa, turns out to be of the order of months. This time is determined by two considerations, the half-lives of the beta-decaying nuclei and their abundances. In the case of Co^{57} the lifetime is long but the abundance fairly high, as has been mentioned. In the case of Mn^{56} the lifetime is short, but the abundance comparatively low. For the values of ρ, T of interest in the present section the free neutron density n_n is too low for free neutron decay to be of much importance in establishing (4).

Reference to Fig. II,4 shows that the time-scale becomes too short to maintain (4) as the explosion point is approached. Equation (5) must then be used, the constant on the right-hand side being obtained by the values of $n(A,Z)$, n_p, n_n that are operative in (4) at the time that we switch from (4) to (5). This statement implies a discontinuous transition between (4) and (5). A rigorous treatment would provide for a continuous transition so that in this sense the present arguments are approximate. Throughout the remaining evolution to the explosion point, and in the subsequent cooling, θ remains approximately constant. The basic reason for this is that although complete equilibrium between prompt nuclear reactions is obtained the neutron-proton ratio is determined by the beta-decay rates and the majority of these decays no longer have time to take place. The question now arises as to what this constant value of θ should be. The values of $n(A,Z)$, n_p, n_n at the time (4) ceases to be operative would be sufficient to determine θ, in addition to determining the constant on the right-hand side of (5). The values of $n(A,Z)$, n_p, n_n in question are in turn determined by the values of ρ, T_9 that are operative at the stage at which (4) ceases to hold good. Thus if these values of ρ, T_9 were known, θ could be calculated. Unfortunately the relevant values of ρ and T_9 are not known *a priori*. Only accurate quantitative calculations of stellar evolution can provide this information, and such calculations have not yet been made.

Our method of solution has been to carry out a series of abundance calculations for a network of values of θ and T_9 and to obtain the values of θ and T_9 which give the best fit to the observed relative abundances. To do this it is convenient to determine the abundances of all of the isotopes relative to Fe^{56}, which is shown later to be the most abundant nucleus. The equation for Fe^{56} is obtained by substituting $A=56$, $Z=26$ in (2); equations for the abundances of the other isotopes relative to Fe^{56} are then obtained by eliminating n_n between the equation for Fe^{56} and the general equation (2). We then obtain

$$\log\frac{n(A,Z)}{n(56,26)}=\frac{A-56}{56}\left\{\log\frac{n(56,26)}{\omega(56,26)}-33.77-\tfrac{3}{2}\log T_9-\tfrac{3}{2}\log 56\right\}+\frac{5.04}{T_9}\left\{Q(A,Z)-\frac{A}{56}Q(56,26)\right\}+\log\frac{\omega(A,Z)}{\omega(56,26)}+\theta\left(Z-\frac{26}{56}A\right)+\frac{3}{2}\log\frac{A}{56}. \quad (6)$$

A considerable simplification can now be effected at the expense of only a slight approximation. This rests on two points, first that $A=56$, $Z=26$ is the most abundant nucleus, and second that the application of (6) is confined to values of A not very different from 56, since nuclei with appreciably different A turn out to possess only negligible abundances; the exception to this is obtained when $A=4$, $Z=2$, so that the present approximation cannot be made if it is desired to work out the He^4 abundance. In this approximation $A/56$ is close to unity and the last term on the right-hand side of (6) may be neglected. More important, the quantity $(A-56)/56$ is small compared to unity, a circumstance that enables us to approximate to the factor that multiplies $(A-56)/56$. For values of ρ between 10^4 g/cm³ and 10^9 g/cm³, which covers the range of density of present interest, we write

$$\frac{A-56}{56}\left\{\log\frac{n(56,26)}{\omega(56,26)}-33.77-\tfrac{3}{2}\log T_9-\tfrac{3}{2}\log 56\right\}$$

$$\simeq 0.183(56-A).$$

The coefficient 0.183 is strictly correct for $T_9=3$, $\rho=2\times10^5$ g/cm³, but a change to $\rho=10^7$ g/cm³ only alters the coefficient to 0.153. This approximate term is in any case of comparatively minor importance in its effect on the value of $\log n(A,Z)$. We now have

$$\log\frac{n(A,Z)}{n(56,26)}=0.183(56-A)+\frac{5.04}{T_9}\left[Q(A,Z)-\frac{A}{56}Q(56,26)\right]+\log\frac{\omega(A,Z)}{\omega(56,26)}+\theta(Z-0.464A). \quad (7)$$

Reference to Fig. II,4 shows that the temperature increases up to the explosion point and cools thereafter. Through the increase of temperature the equilibrium abundances will change in accordance with (7). The abundances will also change during the early part of the cooling. But eventually the temperature will fall so low that the nuclear reactions slow up and the isotopes become "frozen." Freezing occurs when (γ,α), (γ,n), and (γ,p) reactions become insufficient to provide a ready supply of alpha particles and free neutrons and protons. Because the thresholds for these reactions differ from one nucleus to another, freezing will not occur simultaneously for all nuclei, and so it will occur over a range of temperature. Thus the particular value of the temperature that we find to give the best fit to the observed abundances should be interpreted as the *mean freezing temperature.*

Calculations of relative abundances have been carried out for values of T ranging from 2.52×10^9 degrees to 7.56×10^9 degrees, and values of θ of 1.5, 2.5, and 3.5. The binding energies $Q(A,Z)$ have been obtained from the tables of Wapstra and Huizenga (Wa55, Hu55), and the weight factors $\omega(A,Z)$ have been estimated from the level diagrams of Way *et al.* (Wa55a). In one case, that of Ni^{58}, an additional binding of 0.84 Mev over that given by Wapstra and Huizenga has been used to bring the Q into agreement with the mass-spectrometer value.

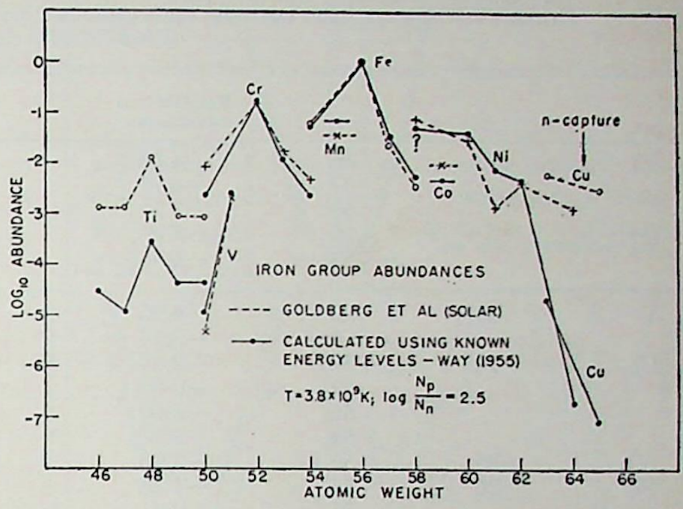


FIG. IV,1. Logarithmic abundance curve for the iron group isotopes relative to the abundance of Fe^{56}. The points connected by dashed lines are the solar abundances determined by Goldberg *et al.* (Go57) using terrestrial isotope ratios to calculate isotope abundances. The points connected by solid lines are calculated from equilibrium theory using partition functions calculated from the energy level data of Way *et al.* (Wa55a) and using a temperature of 3.8×10^9 degrees and a ratio of protons to neutrons equal to $\sim$300.

The best fit is obtained for $T=3.78\times10^9$ degrees and $\theta=2.5$. This value of θ corresponds to a density of the order of 10^5 g/cc. Results of this calculation are shown in Table IV,1 and in Fig. IV,1. To determine the final abundances of the stable isotopes the last reactions, i.e., the freezing reactions, have to be taken into account, together with the beta processes which take place after the freezing reactions have occurred. These reactions are given in column 4 of Table IV,1. The particular freezing reactions which we have taken into account are $V^{49}(\gamma,p)$, $Mn^{53}(\gamma,p)$, $Co^{57}(\gamma,p)$, and $Cu^{63}(\gamma,p)$, all of which have fairly low thresholds of 6–7 Mev and which will freeze at a lower temperature than the (γ,α) and (γ,n) reactions. We have arbitrarily supposed that in these cases half of each of these isotopes have gone to Ti^{48}, Cr^{52}, Fe^{56}, and Ni^{62}, respectively. In columns 6 and 7 of the table and in Fig. IV,1, these calculated abundances relative to Fe^{56} have been compared with the solar element abundances given by Goldberg *et al.* (Go57) taken together with the well-known isotopic abundances. For $50\leq A\leq 62$ the fit is very good indeed, the standard deviation between calculated and observed values being 0.3 in the logarithm in this region. The binding energies are in general uncertain to within 0.2 Mev, and this uncertainty is sufficient to introduce deviations of the same order as those found.

TABLE IV,1. Comparison between $\log_{10}$ observed and calculated abundances in the iron peak (relative to Fe^{56}); $T=3.78\times10^9$ degrees, $\log_{10}n_p/n_n=2.5$.

	Z	A	Equilibrium abundance	Freezing reactions	Final calculated	Observed solar	Observed − calculated
Ti	22	46	−4.53		−4.54	−2.88	+1.66
		47	−4.95		−4.96	−2.89	+2.07
		48	−3.61	$V^{49}(\gamma,p)$	−3.56	−1.92	+1.64
		49	−5.06	$V^{49}(\beta^+)$	−4.37	−3.04	+1.33
		50	−4.36		−4.37	−3.06	+1.31
V	23	50	−4.96		−4.97	−5.31	−0.34
		51	−3.14	$Cr^{51}(\beta^+)$	−2.62	−2.71	−0.09
Cr	24	50	−2.64		−2.65	−2.09	+0.56
		52	−0.80	$Mn^{53}(\gamma,p)$	−0.78	−0.82	−0.04
		53	−2.60	$Mn^{53}(\beta^+)$	−1.93	−1.76	+0.17
		54	−3.30	$Mn^{54}(\beta^+)$	−2.64	−2.32	+0.32
Mn	25	55	−2.06	$Fe^{55}(\beta^+)$	−1.16	−1.44	−0.28
Fe	26	54	−1.26		−1.27	−1.21	+0.06
		56	0	$Co^{57}(\gamma,p)$	0	0	STAN.
		57	−1.94	$Co^{57}(\beta^+)$	−1.49	−1.63	−0.14
		58	−2.36	$Co^{58}(\beta^+)$	−2.27	−2.46	−0.19
Co	27	59	−2.62	$Ni^{59}(\beta^+)$	−2.33	−2.00	+0.33
Ni	28	58	−2.42		−1.31	−1.11	+0.20
		60	−1.35		−1.36	−1.52	−0.16
		61	−2.11		−2.12	−2.85	−0.73
		62	−2.34	$Cu^{63}(\gamma,p)$	−2.35	−2.38	−0.03
		64	−6.86	$Cu^{64}(\beta^+)$	−6.70	−2.88	+3.82
Cu	29	63	−4.36	minus $Cu^{63}(\gamma,p)$	−4.67	−2.17	+2.50
		65	−7.06		−7.07	−2.51	+4.54

A similar uncertainty is also present because of our assumption of a unique freezing temperature.

For $A<50$ and $A>62$ there are gross discrepancies between observed and calculated values and because of these we conclude that other modes of synthesis are demanded for these isotopes. Beyond $A=62$, the nuclei Ni^{64}, Cu^{63}, and Cu^{65} have all been assigned in the appendix to the *s* process and in Sec. VI and in Fig. VI,3, it will be seen that the (σN) products for Cu^{63}, Ni^{64}, Zn^{64}, and Cu^{65} are all comparable ($\sigma N\sim10^4$) as is expected in the first part of the *s*-process chain which runs from $63\leq A\leq 209$. The σN product would not be expected to be constant if these nuclei had been synthesized by the equilibrium process. Below $A=50$, the titanium isotopes have been assigned as follows: Ti^{46} to the *s* process, Ti^{48} to the α process, and Ti^{47}, Ti^{49}, Ti^{50} probably to the *r* process. In this region of A, $46\leq A\leq 50$, where the *s*, α, *e*, and *r* processes all intersect, details of the exact contributions from each process are not easy to determine. In particular, the *r*-process calculations in this region are very approximate (cf. Sec. VII).

V. *s* AND *r* PROCESSES: GENERAL CONSIDERATIONS

The eventual aim is to compare calculated abundances with observed abundances both for the *s* and the *r* processes. Our method of assignment to these processes is described in Sec. II. Before we describe in detail the ways in which the two processes can be handled, it is necessary to digress in order (i) to review briefly the reasons from the standpoint of nuclear stability why all three processes, the *s*, the *r*, and the *p*, are demanded, and (ii) to describe our neutron capture cross sections, σ, which are given in the appendix and are demanded in the discussion of the *s* process alone.

A. "Shielded" and "Shielding" Isobars and the *s*, *r*, *p* Processes

Let us consider nuclei of the same A but different Z. In Fig. V,1 we show the well-known schematic plot of the stability of such nuclei as a function of Z. In this diagram, Z_A is the charge of the most stable isobar at a given A as given by the smooth Weizsäcker mass law (We35). It can occur near an odd or even integral value of Z. The coefficient in the quadratic dependence of the mass $M(A,Z)$ on $(Z-Z_A)^2$ is $B_A/2$. For odd A (top diagram) there is only one parabola and the nucleus with charge Z closest to Z_A is the stable isobar. Those with smaller Z are electron-antineutrino emitters, while those with greater Z capture electrons and emit neutrinos or emit a positron with a neutrino. In a few cases two nuclei near the bottom of the parabola will apparently be stable, one with a very long but detectable half-life or with too long a half-life to be measured. Rb^{87} is an example of this type; it decays to Sr^{87} with a half-life of 4.3×10^{10} years. An odd-A nucleus produced in neutron capture will always decay eventually to the most stable isobar.

For even A there are two mass parabolas. Those nuclei that have odd Z, odd N $(=A-Z)$ lie on a parabola separated from that for nuclei with even Z, even N by the pairing energy, δ. As shown in the two bottom

diagrams, one, two, or three nuclei can be stable, depending on the ratio of δ_A to B_A. Double beta decay is possible for all but the most stable, but this is a very slow process with half-life $>10^{18}$ years. An example of cases where there are three stable isobars is given by Sn^{124}, Te^{124}, and Xe^{124}. On the other hand, there is only one stable nucleus for each even A for $A\leq 34$ and two stable nuclei for $36\leq A\leq 76$. For $A\geq 78$ there may be two or three stable nuclei for each value of A. In a neutron-capture process at a rate rapid compared to negative beta decay (the *r* process), the nuclei produced lie far up the left-hand side of the parabolas shown. At the termination of the process β^- decay leads to the stable isobar with smallest Z. This is the *shielding* isobar, and the other one or two stable isobars with the same A are said to be *shielded*. In a neutron-capture process at a rate slow compared to negative beta decay (the *s* process), the isobar on the left is sometimes produced, that in the middle is sometimes produced, but never the isobar on the right. Those not produced are said to be *by-passed* in the *s* process. It is to produce the isobar on the right, that having the greatest value of Z, that a third process, the *p* process involving the addition of protons on already existing heavy nuclei, is demanded. Details of this process are described in Sec. IX.

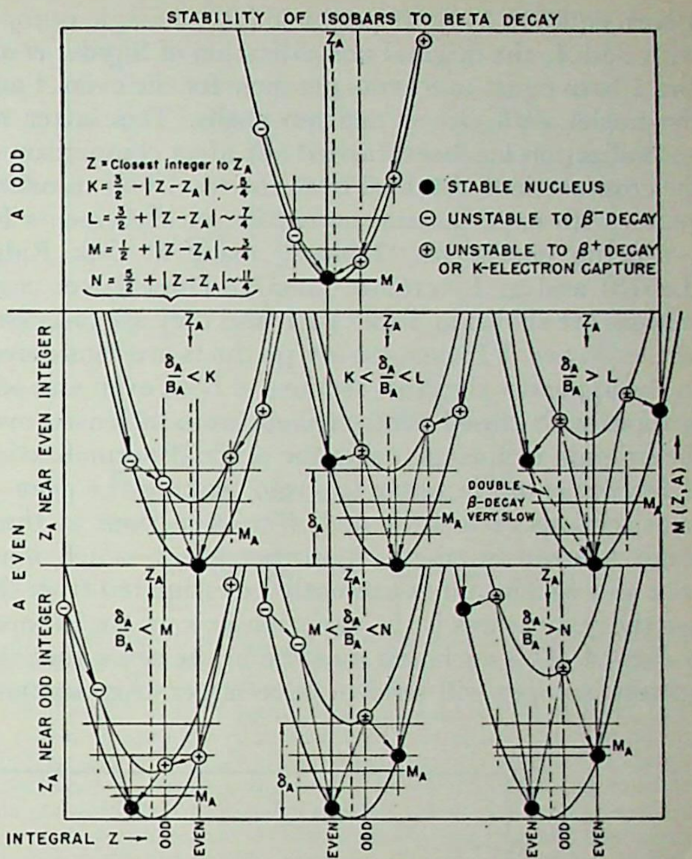


FIG. V,1. The stability of nuclear isobars to beta decay. Z_A is the charge of the most stable isobar at a given A as given by the smooth Weizäcker mass law (We35). It can occur near an odd or even integral value of Z. The coefficient in the quadratic dependence of $M(Z,A)$ on $(Z-Z_A)^2$ is $B_A/2$. For odd A (top diagram) there is only one parabola and the nucleus with charge Z closest to Z_A is the stable isobar. Those with smaller Z are electron-antineutrino emitters while those with greater Z capture electrons and emit neutrinos or emit a positron with a neutrino. In a few cases two nuclei near the bottom of the parabola will apparently be stable, one with a half-life too long to be measured. An odd A nucleus produced in neutron capture will decay eventually to the most stable isobar.

For even A there are two mass parabolas, those nuclei with odd Z, odd N lying on a parabola separated from that for even Z, even N nuclei by the pairing energy δ_A. As can be seen from the two bottom diagrams one, two, or three nuclei can be stable depending on the ratio of δ_A to B_A. Double beta decay is possible for all but the most stable, but this is a very slow process with half-life$>10^{18}$ years. In a neutron-capture process at a rapid rate (*r* process) compared with negative beta decay the nuclei produced lie far up the left-hand side of the parabolas shown. At the termination of the process the negative beta decay leads to the stable isobar with smallest Z. This is the *shielding* isobar and the other one or two are said to be *shielded*. In a neutron-capture process at a slow rate (*s* process) compared with negative beta decay, the isobar on the left is sometimes produced, that in the middle is sometimes produced, but the isobar on the right is never produced. Those not produced are said to be *by-passed* in the *s* process.

B. Neutron-Capture Cross Sections

In the third column of the appendix we give either the neutron-capture cross sections or, in the case of unstable nuclei, the half-lives against beta decay. The cross sections (σ) which are given in millibarns (10^{-27} cm^2) are mostly cross sections for neutron capture with emission of gamma radiation (n,γ). For small A a few cross sections for other reactions initiated by neutrons, in which there is no increase in A [(n,p) reactions] or a decrease in A [(n,α) reactions], are given. All of the values are for a neutron energy of 15 kev which is reasonable if the *s* process is taking place in the interiors of red-giant stars at $T=1-2\times10^8$ degrees. To obtain the cross sections several sources were used:

(i) Many of the σ's for the lighter nuclei have been obtained from the Brookhaven tables (Hu55a). Others have been obtained from the thermal cross sections (1/40 ev) by extrapolation, using a 1/v law, where this law is thought to apply (designated 0.1% of thermal in the appendix). Where σ does not drop as rapidly as 1/v in the kilovolt region we have taken 1% of thermal (designated this way in the appendix) to be a reasonably accurate value. In a few cases the σ's have been marked extr. meaning extrapolated from nearby experimental values, or exp. meaning experimentally measured.

(ii) The majority of the remaining σ's have been obtained from the pile reactivity measurements made by Snyder *et al.* (Sn55), in which reactivity changes due to different materials are measured. Since the reactivity measure is given in units of cents per mole, normalization has been carried out by using the $B^{10}(n,\alpha)$ reaction which has a cross section of 1 barn at 15 kev. No reactivity measures have been given for isotopes with $Z<17$, and so we have made a reasonable extrapolation of the reactivity curves of Snyder *et al.* using the independently measured cross section for Al^{27} of 5.2 mb. The normalization by means of the $B^{10}(n,\alpha)$ reaction for isotopes with $Z>17$ has been carried out by putting

1 barn equal to 24 cents per mole for nonmagic isotopes with odd A, the original normalization of Snyder *et al.*, and 1 barn equal to 8 cents per mole for the even A and the nuclei with closed neutron shells. This latter renormalization has been carried out after comparison of the cross sections obtained from the reactivity measurements with experimental activation measurements for a neutron energy of "25 kev" made at Oak Ridge (La57b) and at Livermore (Bo57). These latter cross sections are shown in Table VI,1, and they are discussed further in Sec. VI when the σN products are considered. In the appendix the cross sections of both even and odd isotopes with closed neutron shells were increased over the original estimates, using the original normalization of Snyder *et al.*, in order to avoid making the even A cross sections exceed the odd A cross sections at these magic-neutron numbers. A further effect which must arise and which cannot presently be estimated from the reactivity measures is the variation in σ within isotopes of even A. This variation must be in the sense that the lightest isotopes will tend to have larger cross sections than the heavier ones. Trends such as these, which may affect the run of the σN products, are discussed further in Sec. VI.

The reason for differentiating between σ's for odd and even A instead of differentiating between nuclei with odd and even numbers of neutrons is the following. The results of Hurwitz and Bethe (Hu51) show that neutron capture occurs through a compound nuclear state whose effective excitation should be measured without regard for special pairing effects in the ground state of this nucleus. Thus only odd-even effects in the ground state of the target nucleus are relevant. The even A-even Z-even N target nuclei capture neutrons with less excitation and thus smaller σ's than the odd A target nuclei. In particular, the value of σ for odd A nuclei is the same for the odd Z-even N case as for the even Z-odd N case. Snyder *et al.*, have taken this effect into account, and in fact their plotted curves give a large odd-even effect. The difference between the normalizations that we have used for odd and even A is in order to *diminish* the odd-even effect

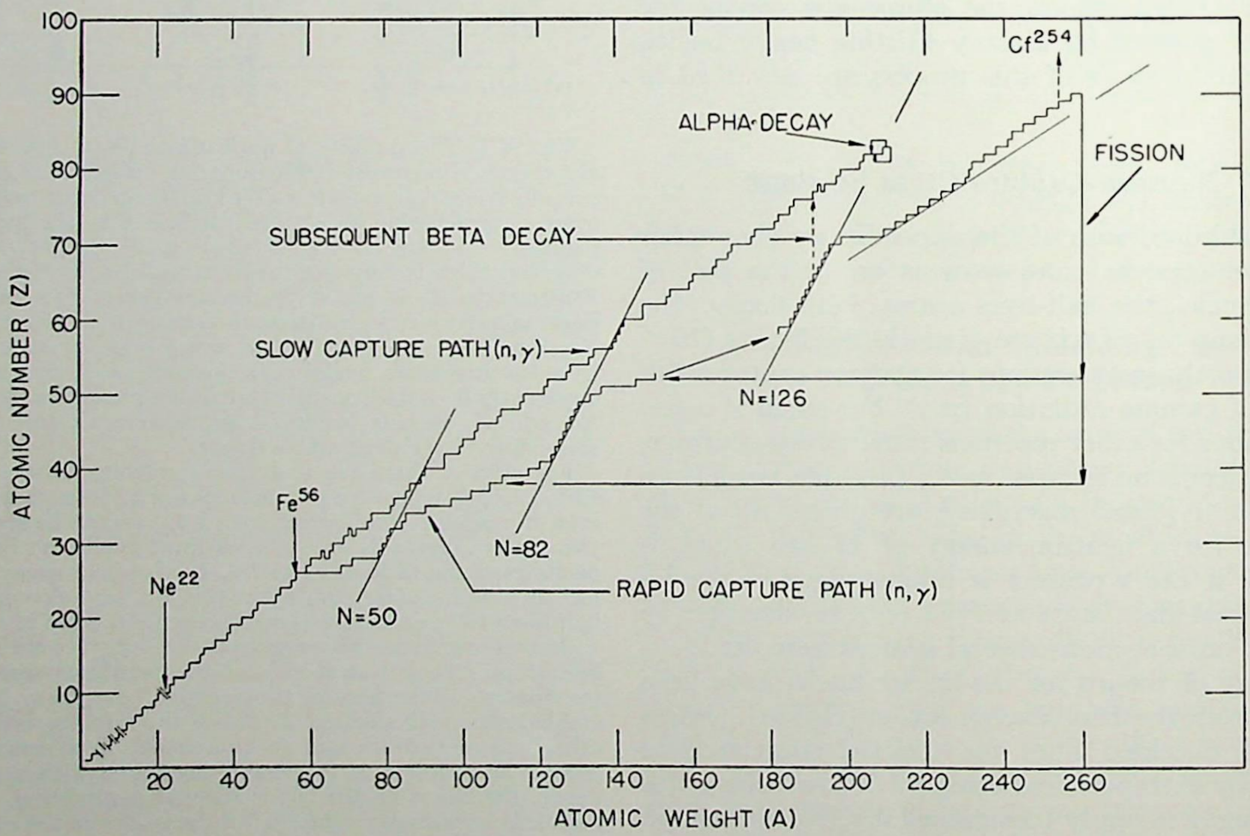


FIG. V,2. The neutron capture paths of the s process and the r process. Ne^{22}, as a typical light nucleus, and Fe^{56} as the most abundant member of the iron group are indicated as "seed" nuclei at which synthesis begins. The capture path of the s process (slow compared with β^- decay) wanders along the stability line in the Z, A plane and finally terminates in alpha decay above bismuth. The r process (rapid compared to β^- decay) is caused by an intense neutron flux which drives the nuclear matter to the neutron-rich side of the stability line. For a temperature of 10^9 degrees the nuclei add neutrons until the neutron binding energy decreases to 2 Mev. At these "waiting points" β decay must occur before further neutrons can be added. Note the "staircase" effects at neutron magic numbers, N, which is due to the difficulty, at this point, of adding the $(N+1)$th neutron. When the path thus moves closer to the stability line the β^- decay lifetime at the waiting points increases and the material builds to an excess abundance. Thus this effect at $N=50$, 82, 126 leads to relatively sharp peaks in the abundance curve. These peaks occur with a smaller Z and A (proton-poor) than is associated with the stable nuclei with magic N and they are thus displaced 6 to 10 units in A below the magic peaks produced in the s process. The r process path is terminated by neutron-induced fission at $A\sim260$ and the nuclear matter is fed back into the process at $A\sim108$ and $A\sim146$. Thus cycling and true steady flow occurs above $A\sim108$.

given by their curves, because we are dealing with neutrons of higher energy than they were using (15 kev instead of $\lesssim 1$ kev).

(iii) For the few cross sections for which estimates are not available either from the thermal measurements or the reactivity measurements, we have interpolated from a smooth curve through the reactivity measurements for $A>40$ and the thermal measurements for $A<20$. In this interpolation we have maintained a ratio of ~ 3 for $\sigma(\text{odd } A)/\sigma(\text{even } A)$ with a minimum at the isotopes with a magic number of 14 neutrons.

C. General Dynamics of the *s* and *r* Processes

In the buildup of nuclei by the *s* and the *r* processes the reactions which govern both the rate of flow and the track followed in the (A,Z) plane are the (n,γ) and (γ,n) reactions, beta decay, and, at the ends of the tracks, alpha decay in the case of the *s* process and neutron-induced fission in the case of the *r* process. We denote the rates of the (n,γ), (γ,n) and beta process as λ_n, λ_γ, λ_β, where

$$\begin{aligned} \lambda_n &= 1/\tau_n = \sigma_n v_n n_n, \\ \lambda_\beta &= 1/\tau_\beta = \text{const}/W_\beta{}^5, \\ \lambda_\gamma &= 1/\tau_\gamma = \sigma_\gamma c n_\gamma, \end{aligned} \tag{8}$$

and σ_n and σ_γ are the cross sections for the (n,γ) and (γ,n) reactions, respectively; v_n and n_n are the velocity and density of neutrons responsible for the (n,γ) reactions; n_γ is the density of γ radiation; and W_β is the beta-decay energy.

The general equation for the buildup of nuclei in the *s*-process is then

$$\begin{aligned} dn(A,Z)/dt = {} & \lambda_n(A-1,\ Z)n(A-1,\ Z) - \lambda_n(A,Z)n(A,Z) \\ & + \lambda_\beta(A,\ Z-1)n(A,\ Z-1) - \lambda_\beta(A,Z)n(A,Z) \\ & + \text{termination terms due to alpha} \\ & \qquad \text{decay at } A>209. \end{aligned} \tag{9}$$

The general equation for the *r* process is

$$\begin{aligned} dn(A,Z)/dt = {} & \lambda_n(A-1,\ Z)n(A-1,\ Z) - \lambda_n(A,Z)n(A,Z) \\ & + \lambda_\beta(A,\ Z-1)n(A,\ Z-1) - \lambda_\beta(A,Z)n(A,Z) \\ & + \lambda_\gamma(A+1,\ Z)n(A+1,\ Z) - \lambda_\gamma(A,Z)n(A,Z) \\ & + \text{termination terms due to fission} \\ & \qquad \text{for } A \gtrsim 260. \end{aligned} \tag{10}$$

For the *s* process, we have, in general,

$$\lambda_n < \lambda_\beta (\tau_n > \tau_\beta).$$

For the *r* process, we have

$$\lambda_n > \lambda_\beta (\tau_n < \tau_\beta).$$

As long as $\lambda_n > \lambda_\beta$ buildup continues with Z constant and λ_n continuously decreases until $\lambda_n(A,Z) \approx \lambda_\gamma(A+1,\ Z)$. At this point no further buildup can take place until

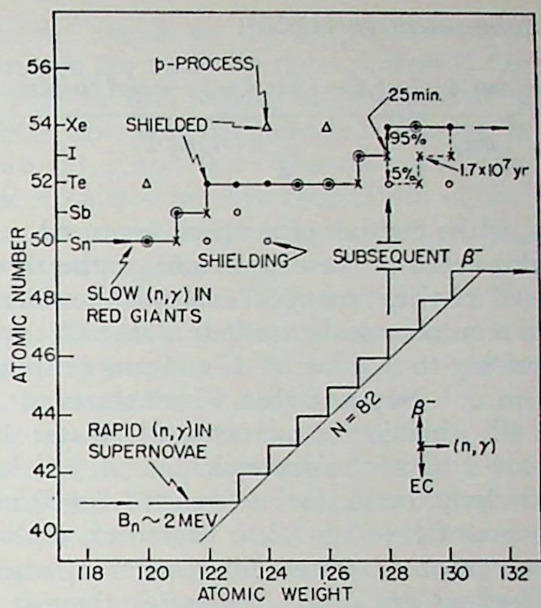


FIG. V,3. Detail of Fig. V,2 showing the *s*-process and *r*-process paths from $A=120$ to 130. Note that the *s* process builds through one stable isobar, except when branching in the beta decay occurs as at $A=128$. The nuclei not built are said to be *by-passed*. In the *r* process the capture path involves those neutron-rich nuclei with binding energy ~ 2 Mev. The path runs parallel to the stability line except at a magic number such as $N=82$, when it turns toward the stability line producing waiting points with greater β^- lifetimes. The subsequent β^- decay in the freezing of the *r* process produces only the *shielding* and not the *shielded* isobars. It will be clear from the diagram that certain nuclei are not produced in either the *s* or *r* processes. These are very rare in abundance as might be expected, and are produced by a secondary process, the *p* process, which is described in Sec. IX.

beta decay occurs, thereby increasing Z. The effective rate of neutron addition at this point is such that

$$[\lambda_n(A,Z) - \lambda_\gamma(A+1,\ Z)] < \lambda_\beta(A,Z).$$

The tracks of both the *s* and the *r* processes in the (A,Z) plane are shown in Figs. V,2 and V,3, and were obtained by methods described in the following sections.

VI. DETAILS OF THE *s* PROCESS

Excluding the alpha-decay terms, and remembering that $\lambda_n < \lambda_\beta$, so that the beta decay takes place in a time short compared with neutron capture, we have from (9)

$$\frac{dn(A)}{dt} = \lambda_n(A-1)n(A-1) - \lambda_n(A)n(A). \tag{11}$$

The large λ_β terms in (9) serve only to produce the most stable Z at each A. Equations (11) determine for the synthesis of the *s*-process nuclei as a function of time. In order to use this type of equation for calculation it is convenient to use the capture cross sections $\sigma(A,Z)$ and to use, instead of t, the number of neutrons injected per standard nucleus, since this is a number which can be calculated from the abundances of the nuclei in the neutron-producing reactions. Thus Eqs.

(11) can be rewritten (Fo55),

$$\frac{dn(A)}{dn_c}=\frac{\sigma(A-1)n(A-1)-\sigma(A)n(A)}{\sum_A \sigma(A)n(A)}, \quad (12)$$

where n_c is the number of neutrons captured.

The properties of these equations can be thought of in terms of a hydrodynamical analogy. Consider a river bed with a number of depressions in it, each depression corresponding to a value of A, and proportional in its volume to σ^{-1}. Suppose that water starts at A_i and flows in the direction of increasing A. Water does not reach (A_i+2) until the depression at (A_i+1) has been filled. Similarly, water does not reach (A_i+3) until the depression at (A_i+2) is filled. Moreover, if there is a final reservoir at A_f, water only reaches A_f when all of the depressions are filled. Thereafter there is steady flow, water going from A_i to A_f without change at intermediate points. This analogy is not complete since in the hydrodynamic case there is no flow at all beyond a particular depression until this depression has been filled, whereas in the nuclear case there is always a finite probability that some flow occurs.

The saturation or steady-flow abundance of a particular isotope is reached when $dn(A)/dn_c=0$, i.e., when

$$\frac{n(A)}{n(A-1)}=\frac{\sigma(A-1)}{\sigma(A)}.$$

If a sufficient number of neutrons is available, the last of the isotopes in the s chain will achieve the full value of $n(A)$ given by this equation and we shall have

$$n(A)\sigma(A)=\text{constant}$$

over the range of isotopes built by the s process, apart from those immediately at the beginning of the chain in which the σN product will decrease unless these nuclei are continuously augmented during the s process. However, if an insufficient number of neutrons is made available, we find that a plot of the $n(A)\sigma(A)$ product is a monotonic function of A, smoothly decreasing as A increases.

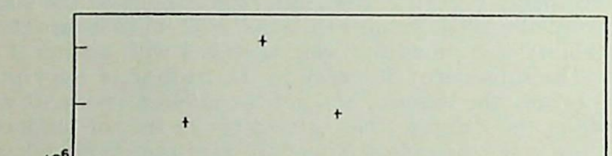


FIG. VI,1. Plot of the abundances of the isotopes times their neutron-capture cross sections (σN products) for $22\leq A\leq 50$. This plot should be compared with an earlier one given by Fowler *et al.* (Fo55). With the exception of those for $A>46$, the points marked by filled circles represent isotopes made predominantly by the s process. Those marked by crosses lie above the other points because of the large abundances of these isotopes, and thus we attribute the synthesis of these isotopes to the α process. The curves are those calculated by R. Tuttle for different numbers of neutrons made available per Ne^{22} nucleus. It can be seen that the best fit to the observed points is given by the curve for $n_c/Ne^{22}=2.1$. The departure of the calculated curve from the plotted points beyond $A\simeq 45$ suggests that other processes must be responsible for synthesis in this region.

There are two separate regions in the abundance curve where a considerable proportion of the isotopes are made by the s process. The region $22<A<50$ has been discussed previously (Fo55) and it was found that in this region steady flow was not attained, but that with a supply of neutrons such that $n_c/n(Ne^{22})=2.8$, the abundances of the isotopes built in the s chain up to Ti^{50} could be reasonably well reproduced. However, some modification of these results can now be made with the cross sections which are given in the appendix. A new plot of σN against A is shown in Fig. VI,1. The effect of the improved cross sections is to show very clearly that σN is truly a very smooth function of A for all of the isotopes built predominantly by the s process in the range of atomic weight $23<A<46$, thus bearing out the correctness of our assignment of these nuclei. One point of special interest is the case of scandium. This has only a single isotope, Sc^{45}, and is made according to the appendix by the s process. However, when the σN product is plotted by using the abundance given by Suess and Urey (Su56) the point falls below the curve delineated by the surrounding points. On the other hand, the solar abundance determined by Goldberg *et al.* (Go57) is larger than that given by Suess and Urey by a factor $\sim$20, and this σN product (which is the one plotted) lies above the curve. This strongly suggests to us that in the case of scandium the true abundance lies between the solar abundance and the estimate from terrestrial and meteoritic data. The points which lie above the curve, designated by crosses instead of filled circles, are due to the nuclei which have $A=24$, 28, 32, 36, 40, 44, and 48. Their values of σN are higher than those of their neighbors because their abundances are far larger. These are the nuclei which are built predominantly by the α process. The neutron-

capture cross sections, σ, clearly have no significance in their case.

Calculations similar to those carried out previously (Fo55) using the new cross section estimates have been carried out by Tuttle (private communication) and curves showing these results are also shown in Fig. VI,1. The best fit is obtained for $n_c/\mathrm{Ne}^{22}=2.1$, as compared with the value of $n_c/\mathrm{Ne}^{22}=2.8$ obtained in the original calculations. This new calculated curve fits the observed σN values only out to $A=43$, whereas in the earlier calculations it appeared that a reasonable fit out to $A=50$ was obtained. Thus, unless the cross section estimates in the region $45<A<50$ are considerably in error, we conclude from these new results that Ti^{47}, Ti^{48}, Ti^{49}, and Ti^{50} are not built predominantly by the *s* process, because their abundances are larger than we should expect on the basis of this theory. As described in Sec. III D, Ti^{48} is built by the α process, and Ti^{47}, Ti^{49}, and Ti^{50} have been tentatively ascribed to the *r* process. This region of A is the region where the α, *s*, *e*, and *r* processes all intersect, a point which was emphasized at the end of Sec. IV. Ti^{46} is assigned partly to the *s* process and partly to the *e* process but its atomic abundance is higher than expected on this basis.

The second region in the abundance curve where many isotopes are made by the *s* process is the region $63\leq A\leq 209$. In this region, as described in Sec. II, assignments are more complicated because in many cases both the *s* and the *r* process contribute to the synthesis of an isotope. The correctness of our assignments can be tested by plotting both the σN products and also by plotting N and $1/\sigma$ as a function of A. These plots are shown in Figs. VI,2 and VI,3. In Fig. VI,2, which shows the plots of N and $1/\sigma$ for isotopes of odd A only, the major peaks in both N and $1/\sigma$ at magic neutron numbers $N=50$ and 82, the approach towards the peak at the double magic numbers $Z=82$ and $N=126$, and the minor peaks at $Z=50$ and $A=180$ are well defined. The run of the points throughout these curves shows that there are no major inconsistencies in our assignments. For a more detailed comparison we use the plot of σN against A (Fig. VI,3). In this plot isotopes made solely by the *s* process are distinguished from those in which the *s* process and the *r* process both contribute roughly equal amounts or in which the *s* process dominates. In the latter two cases the σN product has been corrected by assuming that either $\frac{1}{2}$ or $\frac{2}{3}$ of the abundance is due to the *s* process. In cases

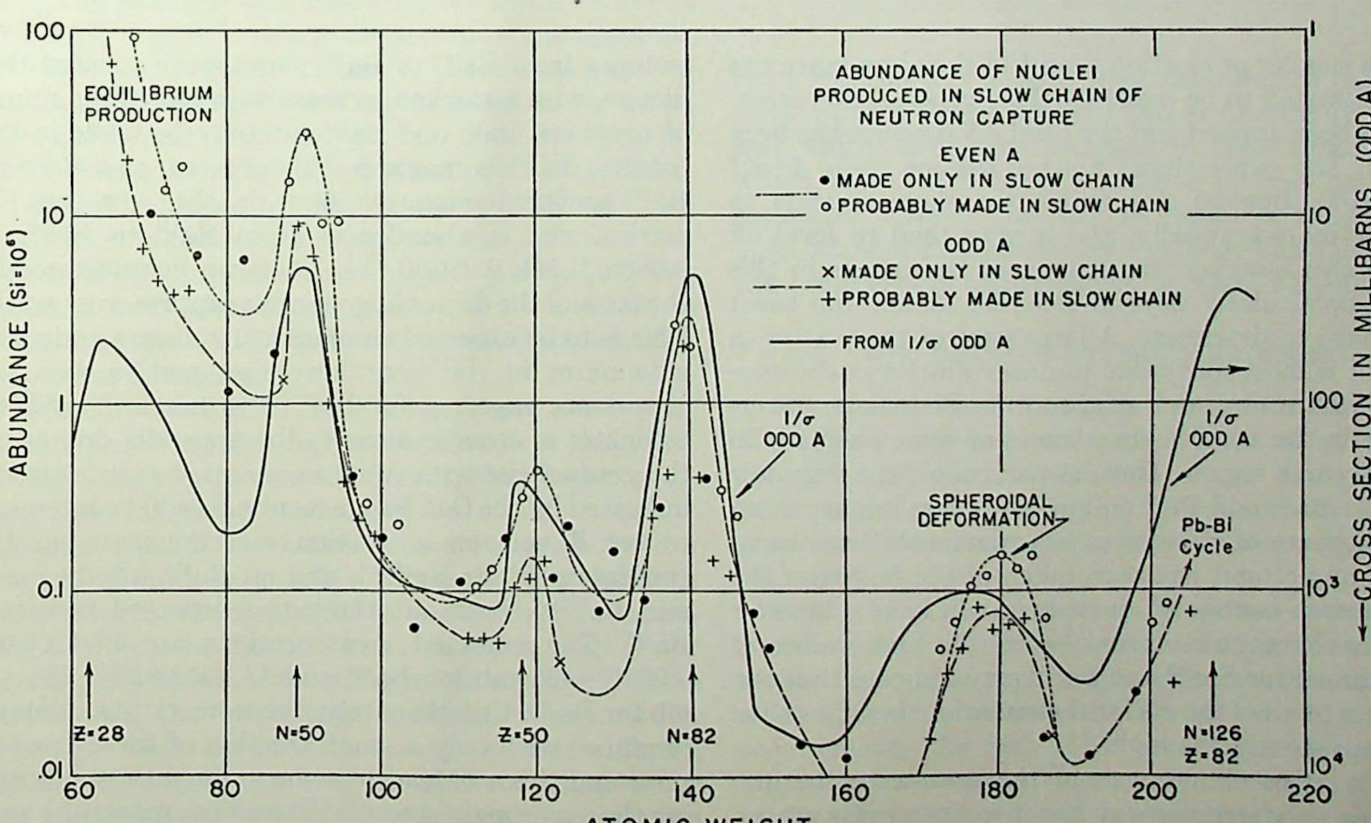


FIG. VI,2. The filled circles and diagonal crosses represent the abundances of even-A and odd-A isotopes, respectively, in the range of atomic weight $63\leq A\leq 209$ which are made only in the *s* process; open circles and vertical crosses represent abundances of even- and odd-A isotopes, respectively, which are probably made in the *s* process. The abundance scale is given on the left-hand ordinate. Two curves are drawn schematically through the even- and odd-A points separately to show the trends of these abundances. The continuous curve which is also given is a plot of σ^{-1}, the reciprocal of the neutron capture cross section for isotopes of odd A, this scale being given by the right-hand ordinate. It will be seen that the peaks at the neutron magic numbers $N=50$ and 82, the proton magic number at $Z=50$, and the rise to the doubly-magic peak at $N=126$, $Z=82$, together with the broader rise due to the spheroidal deformation effect near $A=180$, are all shown by these curves, reflecting the effect that in local regions the σN product is constant, though over the whole range of A which is displayed, the σN product shows considerable variation (cf. Fig. VI,3).

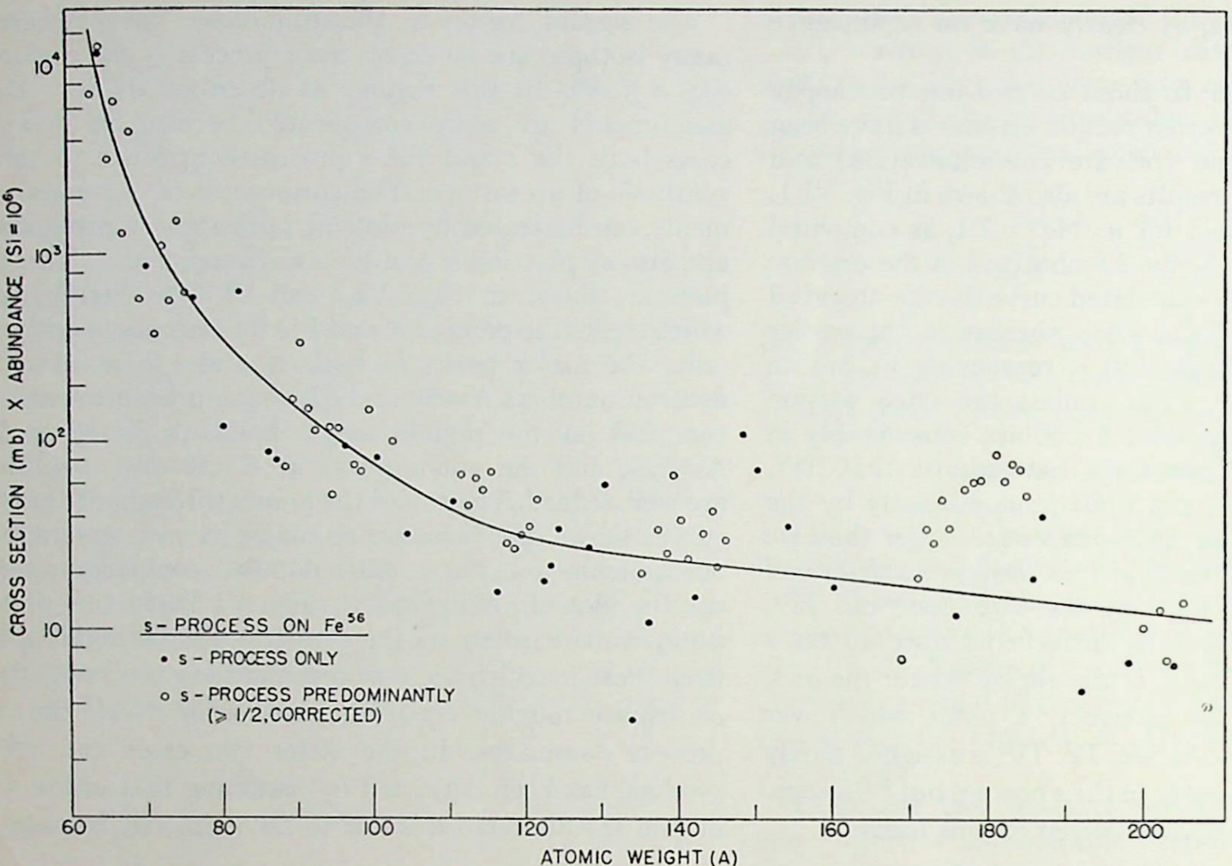


FIG. VI,3. Plot of the abundances of the isotopes times their neutron-capture cross sections (σN products) for $63 \leq A \leq 209$. The points marked by filled circles represent isotopes made by the *s* process only, while those points marked by open circles represent isotopes made by the *s* process predominantly; in cases in which it has been estimated that a certain fraction of the abundance is made by the *s* process (perhaps $\frac{1}{2}$ or $\frac{2}{3}$) this fraction was used in computing σN. The curve which is drawn is purely schematic.

where a smaller proportion than $\frac{1}{3}$ of the abundance has been estimated to be made by the *r* process, no correction has been applied and the total σN product has been plotted. The curve shows a steep decline from $A \sim 63$ to $A \sim 100$. Beyond this point, the slope appears to decrease more gradually, and it may tend to level off completely; however, the scatter of the points in this region, particularly beyond $A \sim 170$, makes the exact trend hard to determine. A large part of the scatter in this plot is most likely due to uncertainties in the estimated values of σ and also, to a lesser extent, uncertainties in the relative abundances of some elements in the rare earth region. Thus, in particular, the very low points at Ba^{134} and Pt^{192} are probably due in part to an underestimate of σ, since as has previously been mentioned, no account has been taken of the fact that the lightest even isotope of an element will have a larger σ than heavier even isotopes. Again, the high values of σN obtained for Sm^{148} and Sm^{150} may indicate that our estimates of all of the σ's for the samarium isotopes have been consistently too high.

The relative abundances of the isotopes of barium show in a very striking way how the *s*-process operates. The light isotopes Ba^{130} and Ba^{132} cannot be made on the *s* or *r* processes and are indeed very rare with atomic abundances 0.0037 and 0.00356, respectively. The fact that they are almost equal can be well understood on the basis of their production in the *p* process (cf. Sec. IX). They result from the decay of positron-emitting progenitors at 130 and 132, made in equal abundance in the *p* process. On the other hand, the isotopes from Ba^{134} to Ba^{138} show much greater abundances, with a marked increase with increasing number of neutrons. Ba^{134} and Ba^{136} can only be made in the *s* process, Ba^{138} is magic in this process, and Ba^{135} and Ba^{137} are predominantly made in the *s* process. The marked rise in abundances from Ba^{134} to Ba^{138}, i.e., 0.0886, 0.241, 0.286, 0.414, 2.622 can be understood on the basis of the decreasing neutron capture cross section. This is to be expected as more neutrons are added, and culminates in the very low cross section and large abundance expected for Ba^{138} with magic $N=82$. Our estimates of cross sections in the appendix do not take the gradual rise with N into account, but it is strongly indicated by the Oak Ridge results. It will be interesting to see if neutron activation measurements on Ba^{134} (producing 29 hr Ba^{135m}), and on Ba^{136} (producing 2.6 min Ba^{137m}), bear out the ratios expected relative to Ba^{138}. The expected cross sections are $11.4 \times 2.622/0.0886 = 336$ mb for Ba^{134} and $11.4 \times 2.622/0.286 = 109$ mb for Ba^{136}. Unfortunately, the isomeric states may be produced with only a small fraction of these cross sections and this will lead to some ambiguity in interpreting the experimental result. Unambiguous results would be obtained by measuring the total absorption cross sections, but this is very difficult experimentally. It is our view, however, that such measurements would serve as a crucial test of the validity of the *s* process.

The steep part of the curve shown in Fig. VI,3, followed by part which tends to level off, suggests that the curve has a composite character and that the abundances in the solar system of *s*-process isotopes

have been produced in more than one stellar synthesis process. Thus, the simplest possibility is that one process is responsible for the part of the curve in the range $63<A\lesssim 100$ and that this part is similar in shape to the σN plot for the isotopes in the range $33<A<46$, shown in Fig. VI,1, falling to zero very rapidly beyond $A\sim 100$. This part of the curve shows that there have not been enough neutrons available per Fe^{56} nucleus to build the nuclei to their saturation abundances. Thus, it may be that in this region the $C^{13}(\alpha,n)O^{16}$ reaction has been the neutron source. A second process may be responsible for that portion of the curve beyond $A\sim 100$. In this case, since this latter slope is small or zero, this is strongly suggestive of steady flow being achieved and of all of these nuclei reaching their saturation abundances. The rates at which the neutrons are released by $C^{13}(\alpha,n)O^{16}$ and $Ne^{21}(\alpha,n)Mg^{24}$ are discussed in Sec. III. The time-scales obtained from arguments based on particular isotopes in the *s*-process chain are given in Sec. II, where it was shown that these time-scales were compatible with the result that the σN product falls steeply to $A\sim 100$, implying a paucity of neutrons, and then levels off, the abundances reaching their saturation values. Thus at the end of the *s*-process chain we expect that cycling among the lead isotopes, following the alpha decays at Po^{211}, Bi^{210}, and Bi^{211} which are shown in the appendix and illustrated in Figs. V,2 and V,3 has taken place, so that the lead isotopes which are included in the cycle, Pb^{206}, Pb^{207}, and Pb^{208}, have been built up to a far greater extent than Pb^{204} which is not in the cycle. Some lead is also produced by the *r* process, and if this contribution is taken into account we can calculate the amount of cycling which has taken place, and also predict a total abundance of lead. This question is discussed in Sec. VIII, following detailed explanation of *r*-process dynamics.

If this curve in Fig. VI,3 is a composite of two curves, it is also clear that the degrees of dilution of the abundances in the two parts are different, since larger over-abundance ratios are demanded to obtain complete saturation than if saturation is not achieved. Such a situation is entirely possible, since the two different processes might have occurred in two different red-giant stars, in which the dilution was a function both of the original composition, the mixing to the surface, and the ejection into the interstellar medium. Alternatively, they might have occurred in a single star at different evolutionary stages in which the structures and hence these same parameters would be changed. To study this question further it is important to investigate the composition of a star in which it is believed that the *s* process has recently been occurring. The *S*-type stars and the so-called Ba II stars fall into this category, and the results of such an investigation are described in Sec. XI D.

Finally, in Table VI,1 we show the σN products for just those isotopes for which experimental measures of

TABLE VI,1. "25 kev" cross section (La57b, Bo57) ×abundance (Su56).

	σ(mb)	$N(Si=10^6)$	σN	Assignment
I. Synthesized on Ne^{22}				
Na^{23}	1	4.38×10^4	4.38×10^4	H burning, s
Mg^{26}	<14	1.00×10^5	$<1.4\times 10^6$	s(m)
Al^{27}	1.4	9.48×10^4	1.3×10^5	s(m)
Cl^{37}	1.1	2180	2.4×10^3	s(m)
K^{41}	19	219	4.2×10^3	s
II. Synthesized on Fe^{56}				
Cu^{63}	116	146	1.7×10^4	s
Ni^{64}	8	318(0.61)	1.6×10^3	s
Cu^{65}	46	66	3.0×10^3	s
Ga^{71}	68	4.54	3.1×10^2	s
Ge^{74}	54	18.65	1.0×10^3	s
As^{75}	590	4.0(2/3)	1.6×10^3	sr
Y^{89}	25	8.9	2.3×10^2	s(m)
Zr^{94}	243	9.48	2.3×10^3	s
Mo^{98}	209	0.581(1/2)	60	s=r
Ru^{102}	386	0.467(1/2)	90	s=r
In^{115}	805	0.105(1/2)	42	s=r
Sb^{121}	950	0.141	130	s
Ba^{138}	11.4	2.622	30	s(m)
La^{139}	50	2.00	100	s(m)
Ce^{140}	31	2.00	62	s(m)
Pr^{141}	546	0.40	218	s(m)
Hf^{180}	441	0.155	68	s
Re^{185}	2650	0.05(1/2)	66	s=r(m′)
Hg^{202}	57	0.0844(2/3)	3.2	sr
Bi^{209}	1.4	0.144(?)	0.2(?)	{s(m)-cycle {r-decay

cross sections at "25-kev" neutron energies have been made at Oak Ridge (La57b) and at Livermore (Bo57), and which have been assigned to the *s* process or in which the contributions of the *r* process have been estimated. Five of the isotopes lie in the range $23\leq A\leq 46$, while the remainder have $A\geq 63$. This table shows the same effects as are shown in Figs. VI,1 and VI,3.

VII. DETAILS OF THE *r* PROCESS

The essential feature of the *r* process is that a large flux of neutrons becomes available in a short time interval for addition to elements of the iron group, or perhaps, in cases where the abundances in the iron group are abnormally small, for addition to light nuclei such as Ne^{22}.

From the point of view of the present section, the precise source of the neutrons is not important; any source capable of supplying a large neutron flux on a short time-scale of order 10–100 sec, would meet the requirements. In Sec. III F we discussed the possibility that a neutron density of $10^{24}/cm^3$ and a flux of $4\times 10^{32}/cm^3$ sec might be produced in supernova envelopes at temperatures $\sim 10^9$ degrees. We assume these conditions in the following discussion of the path of the *r* process and of the abundances produced by it. Fong (Fo57), using a hydrodynamical analogy, has previously calculated abundances produced by an *r* process. Our method is somewhat different in that we include the (γ,n) as well as (n,γ) and β^- decay in our calculations.

A. Path of the *r* Process

A nucleus of fixed Z cannot add neutrons indefinitely even in the presence of an intense neutron flux. The energy of binding of each successive neutron becomes progressively weaker as more and more neutrons are added, until ultimately the binding falls to zero. This sets an upper limit to neutron addition at fixed Z. Moreover, not even this limit will be attained because of the incidence of (γ,n) reactions at the temperature necessary to produce the neutrons by charged particle interactions. The effect of (γ,n) reactions can readily be understood from statistical considerations. Write $n(A,Z)$ and n_n for the number densities of nuclei A, Z and of neutrons respectively, and let $Q_n(A,Z)$ be the neutron binding *to* the nucleus $X(A,Z)$ in the reaction,

$$X(A,Z)+n\rightleftharpoons X(A+1,Z)+\gamma+Q_n(A,Z),$$

such that

$$Q_n(A,Z)=B_n(A+1,Z)$$
$$=c^2[M(A,Z)+M_n-M(A+1,Z)]. \quad (13)$$

Note that Q_n for nucleus A, Z is equal to the neutron binding energy B_n (taken positive) *in* nucleus $A+1$, Z. With T in units of 10^9 degrees, Q_n in Mev, and neglecting weight factors of order unity, the statistical balance in this reaction is expressed by

$$\log\frac{n(A+1,Z)}{n(A,Z)}=\log n_n-34.07-\tfrac{3}{2}\log T_9+\frac{5.04}{T_9}Q_n \quad (14)$$

with all logarithms to the base 10. For neutron addition to be highly effective we require $n(A+1,Z)/n(A,Z)\gg 1$. When $n(A+1,Z)/n(A,Z)\sim 1$, neutron addition is less effective. When $n(A+1,Z)/n(A,Z)\ll 1$, neutrons are scarcely added at all. As an approximation, adequate for the present discussion, we may consider that neutrons are added or not added according to whether $n(A+1,Z)/n(A,Z)\geq 1$ is satisfied or not, i.e., according to whether

$$Q_n\geq\frac{T_9}{5.04}(34.07+\tfrac{3}{2}\log T_9-\log n_n) \quad (15)$$

is satisfied or not. With $T_9\sim 1$ and $n_n\sim 10^{24}$ as found in Sec. III F, we have

$$Q_n\gtrsim 2 \text{ Mev}.$$

Larger Q_n values would be required at higher temperatures or at lower neutron densities. $n(A+1,Z)/n(A,Z)$ is very sensitive to small changes in Q_n or T_9. Under the conditions discussed above $5.04Q_n/T_9\approx 10$. Hence a ten percent change in Q_n, namely by 0.2 Mev, or a ten percent change in T_9, namely by 10^8 degrees, will change $n(A+1,Z)/N(A,Z)$, by a factor of ten. Conversely, for given Q_n, T_9, and Z the distribution in abundance will exhibit a sharp peak where $n(A+1,Z)=n(A,Z)$. This peak will be at most two or three units in A in width.

We have said that Q_n decreases as more and more neutrons are added at fixed Z. This statement ignored neutron pairing effects. Thus Q_n is higher for odd $N=A-Z$ by 1.5 to 3 Mev (depending on A) than it is for even $N=A-Z$. This means that instead of a simple monotonic decrease of Q_n with increasing A (Z fixed) we have two decreasing sequences, one for N odd, the other for N even, the two sequences being separated by about 2 Mev in the range of A of interest in this discussion.

Consider the effect of these two sequences on the criterion for neutron addition. Evidently the pairing effect makes it substantially more difficult for Q_n to satisfy the criterion when N is even, and hence neutron addition must always be expected to cease when N is even. Thus if a neutron can be added at N even, then a further neutron will always be added, since an additional energy approximately equal to 2 Mev is then available for binding the second neutron. On the other hand, when a neutron is added at odd N, the situation is reversed—a second neutron would in that case be less strongly bound by $\sim$2 Mev—so that, although the first neutron might satisfy our criterion, the second is a good deal less likely to do so, and eventually, when N becomes large enough, will not in fact do so. Thus we expect that neutron addition will always cease with N at an even value, so that if A is odd Z must be odd, and if A is even Z must be even.

These remarks all refer to a fixed value of Z. Now Z must increase by unity from time to time, since the neutron-rich nuclei at present under consideration are all unstable against negative beta decay, the lifetimes being $\sim$0.01–10 sec. When Z thus increases, the energy of binding of the next neutron increases quite substantially, so that one or more neutrons can again be added. Once again, however, neutron addition will cease at some even N. Neutron addition cannot then proceed until Z increases by a further unit—and so on. We therefore obtain a clear picture of how the nuclei evolve, by repeated neutron addition interspersed with β^- processes, the nuclei always being obliged to "wait" for a β^- process when neutron addition decreases Q_n to the limit allowed by our criterion. The process is shown schematically in Fig. V,2. A detail of this figure for $118<A<132$ is shown in Fig. V,3.

The *s* and *r* processes differ in a very crucial way. In the *s* process, neutrons are made available very slowly so that the neutron density is always low. In such circumstances the addition of neutrons to the nuclei is controlled by their (n,γ) cross sections. That is to say, the various nuclei compete among themselves for the capture of a slow trickle of neutrons, the abundances of the nuclei being governed by the flow equation (11)

$$dn(A)/dt=-\lambda_n(A)n(A)+\lambda_n(A-1)n(A-1),$$

where $n(A)$ is the abundance of that beta-stable isobar at A which occurs in the *s*-process capture path. In the *r* process, on the other hand, we are concerned with a

situation in which there is no neutron shortage, and neutron addition is limited not by (n,γ) cross sections but by (γ,n) competition and by "waiting" for β^- processes to take place. Essentially, we assume that equilibrium is reached between the rapid (n,γ) and (γ,n) processes, *viz.*, $n, \gamma \rightleftharpoons \gamma, n$, and that the slow beta-decay processes constitute a slow leakage from this equilibrium. The corresponding flow equation in the r process, derived from the general equation (10), takes the form

$$dn(Z)/dt = -\lambda_\beta(Z)n(Z) + \lambda_\beta(Z-1)n(Z-1), \quad (16)$$

where $n(Z)$ is the density of that isotope of element Z at which waiting for negative beta decay occurs, i.e., the isotope for which $Q_n \sim 2$ Mev. The quantity $\lambda_\beta(Z)$ is related to the beta-decay mean lifetime, τ_β, and half-life, t_β, by $\lambda_\beta = 1/\tau_\beta = 0.693/t_\beta$. The condition for steady flow in the r process is

$$\lambda_\beta(Z)n(Z) = \lambda_\beta(Z-1)n(Z-1) = \text{const}$$

or

$$n(Z) \propto \lambda_\beta^{-1}(Z) = \tau_\beta(Z).$$

This is the analog of the steady-flow condition in the s process, namely,

$$\lambda_n(A)n(A) = \lambda_n(A-1)n(A-1) = \text{const}$$

or

$$n(A) \propto \lambda_n^{-1}(A).$$

Just as the abundances of the s nuclei suggest that steady flow has occurred in the s process, so the abundances of the r nuclei suggest that steady flow has occurred in the r process.

The $n(Z)$ given by the flow equations cannot immediately be interpreted as atomic abundances. At first sight it might seem as if only certain values of A are concerned in the r process, namely the values of A at which the nuclei "wait." In Figs. V,2 or V,3, if nuclei of charge Z wait at A, nuclei of charge $Z+1$ wait at $A+\Delta A$, where ΔA equals the number of neutron captures at $Z+1$ after beta decay at Z. Thus it might appear that only atomic weights A, $A+\Delta A$, etc., are likely to be produced in any appreciable abundances by the r process. This would be the case if there were a unique track of the kind shown in Figs. V,2 and V,3. But this is scarcely likely to be the case, since there will certainly be a spread in A, Z values about this path, simply due to the statistical nature of Eq. (14). Thus a small scatter in the values of atomic weights at which waiting for beta decay occurs is only to be expected. This suggests that we convert the n_Z abundances to n_A abundances by

$$n(A) = n(Z)(\Delta A)^{-1} = n(Z)(dZ/dA), \quad (17)$$

where ΔA is the number of unit steps in A for a single step in Z along the typical track of the form shown in Fig. V,2, and dZ/dA is the slope of the track in the Z, A plane. That is to say, we "spread" the abundances accumulated at the "waiting points" over the intervals of A that occur between the waiting points. This treatment is of course approximate, but it is thought adequate for the present purpose, since the present analysis contains other simplifications that are explained at a later stage. We arbitrarily take the spread from a given waiting point at A to the next higher waiting point at $A+\Delta A$ rather than in the reverse direction. Physically, this corresponds to the assumption that, in freezing, the material accumulated at a waiting point with given A, Z is spread out, as γ radiation dies out, by final neutron captures over an interval in A up to the waiting A for $Z+1$. The equation for the equilibrium between (n,γ) and (γ,n) reactions is much more sensitive to T, (linear dependence), than to n_n, (logarithmic dependence). Thus, as T and n_n decrease, the equilibrium is displaced toward slightly greater A values, representing the capture of the last neutrons.

The solution of the dynamical problem in the r process falls into two parts. The first part consists in a determination of the track shown schematically in Fig. V,2, the second consists in estimating the beta-decay waiting time. The relative abundances given by steady flow then follow immediately from

$$n(A) \propto \frac{\tau_\beta(A,Z)}{\Delta A} = \tau_\beta(A,Z)\frac{dZ}{dA},$$

where $\tau_\beta(A,Z)$ is now the mean beta-decay lifetime at the waiting point, A, for a given Z. A determination of the track demands a precise specification of the criterion for neutron addition. This criterion can be written as $Q_n \geq Q_o$, where Q_o is to be a specified quantity, of order 2 Mev, which depends on n_n, T in accordance with the inequality (15).

The problem of the determination of the track is that of obtaining Q_n as a function of A, Z from nuclear data. Assuming for the moment this to be done, we then have $Q_n(A,Z) \geq Q_o$ as the condition for neutron addition. We now wish to determine the waiting values of A for various Z by using the following criteria.

With Z specified, $N = A - Z$ must be even, and

$$Q_n(A,Z) \geq Q_o,$$
$$Q_n(A+2, Z) < Q_o.$$

In this way the waiting points A, $A+\Delta A$, $\cdots$ corresponding to Z, $Z+1$, $\cdots$ can be found. A further necessary condition is that the waiting points increase monotonically with Z.

We have not found it possible to determine the waiting points solely from the current empirical nuclear data on neutron binding energies. This is to be expected, if it is recalled that one must know the binding energy of the very neutron-rich nuclei along the neutron capture path with considerable precision (~ 0.2 Mev) in order to specify the path precisely. Even the binding energies of the stable nuclei on which careful mass measurements have been made are rarely known to this

precision. Furthermore, the extrapolation to the neutron-rich nuclei involves considerable uncertainty. For these reasons we have been forced to resort to a method of calculating r-process abundances which establishes a few parameters semiempirically on the basis of certain salient features of the abundance curve itself. We now outline a method of calculation which may eventually be capable of yielding a theoretical abundance curve on the basis of nuclear data alone.

First, we consider the determination of $Q(A,Z)$ on the basis of the smooth Weizsäcker atomic mass formula (We35), neglecting shell, pairing and quadrupole deformation effects:

$$M(A,Z)=(A-Z)M_n+ZM_H-\frac{1}{c^2}\left[\alpha A-\beta\frac{(A-2Z)^2}{A}-\gamma A^{\frac{2}{3}}-\epsilon\frac{Z(Z-1)}{A^{\frac{1}{3}}}\right], \quad (18)$$

where M_n and M_H are the masses of the neutron and of the hydrogen atom and α, β, γ, and ϵ are constants in energy units, to be determined empirically. They represent, respectively, volume, isotopic, surface, and Coulomb energy parameters. In the following discussion we use $\alpha=15.3$, $\beta=22.6$, $\gamma=16.7$, and $\epsilon=0.69$, all in Mev, as determined by Fowler, Hornyak, and Cohen (Fo47). Equation (18) can be alternatively written as

$$M(A,Z)=M_A+\frac{1}{2c^2}B_A(Z-Z_A)^2, \quad (18a)$$

where

$$Z_A\simeq\tfrac{1}{2}A\left(1+\frac{\epsilon}{4\beta}A^{\frac{2}{3}}\right)^{-1} \quad (19)$$

is the value of Z at a given mass number, A, for which $M(A,Z)$ has the minimum value, M_A. Z_A is not necessarily integral, and both M_A and Z_A can be determined in terms of the empirical constants by setting $\partial M/\partial Z=0$, keeping A constant. The coefficient B_A appears in the well-known parabolic dependence of $M(A,Z)$ on Z and can be evaluated from

$$B_A=\frac{8\beta}{A}\left[1+\frac{\epsilon}{4\beta}A^{\frac{2}{3}}\right]\simeq\frac{4\beta}{Z_A}. \quad (20)$$

Alternatively, M_A, Z_A, and B_A can be taken as quantities to be determined as empirical functions of A.

Returning to Eq. (18), we can determine $Q_n(A,Z)$ as follows

$$\begin{aligned}Q_n(A,Z)&=c^2[M_n+M(A,Z)-M(A+1,Z)]\\&=\alpha-\beta\left(1-4\frac{Z^2}{A^2}\right)-\tfrac{2}{3}\gamma A^{-\frac{1}{3}}+\frac{\epsilon}{3}\frac{Z(Z-1)}{A^{\frac{4}{3}}}\\&\simeq\alpha-\beta-\tfrac{2}{3}\gamma A^{-\frac{1}{3}}+4\beta\frac{Z^2}{A^2}\left[1+\frac{\epsilon}{12\beta}A^{\frac{2}{3}}\right].\end{aligned} \quad (21)$$

The terms in the second line in Eq. (21) can be identified successively as due to volume, isotopic, surface, and Coulomb effects in atomic masses. Equating Q_n to a specified Q_o in Eq. (21) gives A as a smooth function of Z. The plot of Z against A runs approximately parallel to the stability line, but it is depressed below this line by an amount that depends on the value of Q_o. For $Q_o=2$ Mev the depression amounts to about 7 units in Z at $A\sim100$ and about 11 units in Z at $A\sim200$. If the nuclear masses were really given by the Weizsäcker formula the nuclei would evolve along this line, not completely in a smooth fashion because of the discrete nature of A, but smoothly within one unit of A. There would be no gross deflections of the line, such as are implied in Figs. V,2 and V,3. In addition, the values of τ_β and dZ/dA would also be a smooth function of A, so that the theory would be incapable of explaining the peaks and troughs of the abundance curve for the r nuclei. Evidently an improved expression must be used for $M(A,Z)$ if the theory is to provide this degree of detail. We have not attempted a complete treatment but have been content to approximate in a degree that does not seem to have too serious an impact on the calculated results. A fuller treatment of the form of $M(A,Z)$ is at present being investigated by Dr. Forrest Mozer at the California Institute of Technology, and improved results should be forthcoming from its use.

Our treatment is based on modifying the above expression for $M(A,Z)$ so that we have

$$M(A,Z)=M_W(A,Z)-\frac{1}{c^2}[f(N)+g(Z)], \quad (22)$$

where $M_W(A,Z)$ represents the Weizsäcker expressions given in Eq. (18). That is to say, we subtract a sum of two functions, one of the neutron number and the other of the proton number. Insofar as they are functions of N and Z separately, this procedure takes into account the important effects on nuclear masses of (i) neutron or proton shell structure, (ii) spheroidal quadrupole deformation of partially filled shells, and (iii) pairing of neutrons and pairing of protons. Products of two such functions are not included. The quantities $f(N)$ and $g(Z)$ will be discontinuous functions at magic closed shell numbers for N or Z respectively. The sign is taken negative so that $f(N)$ and $g(Z)$, as positive quantities, decrease the mass and add to the stability of a nucleus.

We now obtain

$$Q_n(A,Z)=\alpha-\beta-\tfrac{2}{3}\gamma A^{-\frac{1}{3}}+4\beta\frac{Z^2}{A^2}\left[1+\frac{\epsilon}{12\beta}A^{\frac{2}{3}}\right]+f'(A-Z), \quad (23)$$

where

$$f'(A-Z)=f'(N)=\frac{df(N)}{dN}=f(A+1-Z)-f(A-Z).$$

Note that $g(Z)$ drops out of Q_n. Care must be exercised in evaluating f' at magic numbers where f is discontinuous. Our criterion $Q_n \geq Q_o$, with Q_o a specified number ~ 2 Mev, now gives a track in the Z, A plane that depends on $f'(A-Z)$.

The capture path is changed considerably on taking $f'(A-Z)$ into account. The plot of Z against A for the r process follows a smooth curve approximately parallel to the stability line if the Weizsäcker mass formula, Eq. (18), is used. The plot of Z against A, shown in Fig. V,2, departs from such a smooth curve, however. Deviations are introduced by the function $f(A-Z)$. Most notably, this function causes the curve to swing upwards towards the stability line whenever $A-Z$ equals a neutron magic number. The physical reason for this is that the energy of binding of the next neutron decreases sharply by about 2 Mev at neutron shell closure. Consider neutron shell closure at $N=A-Z=82$, $A=123$, $Z=41$ (cf. Fig. V,3). The 82nd neutron is added with a binding of 2 Mev, or perhaps somewhat more than this, the value of Z being some 12 units below the stability line at this stage. An 83rd neutron cannot then be added at such a depth below the stability line because of the sharp fall of binding after shell closure. Indeed, an 83rd neutron at $Z=41$ would scarcely be bound at all. Further evolution then depends on a β^- process, which decreases $A-Z$ to 81. An 82nd neutron can then be added, but an 83rd again cannot. A further β^- process must therefore take place, when once again an 82nd neutron can be added. In this way the nucleus climbs a staircase with Z and A both increasing by unity at each step. Eventually, however, the climb up the staircase carries the nucleus close enough to the stability line for an 83rd neutron to be added at a binding energy ~ 2 Mev. This occurs for $Z=49$, $A=131$. Subsequently Z increases by one unit, not for a one-unit increase of A, as on the staircase, but for an increase of some three or more units in A. The differential in A must be an odd integer since waiting occurs only for even N. We refer to the values of A and Z for a magic N at which more than one capture occurs as the "break-through" values at the abundance peaks. Break-through pairs are (82, 32), (131, 49), and (196, 70) at $N=50$, 82, and 126, respectively, and staircases similar to the one at $N=82$ occur at $N=50$ and 126. These staircases are associated with the rising sides of major peaks in the abundance curve for the r process. More subtle effects of proton shells and spheroidal deformations are discussed later.

From the above discussion it would seem that the capture path cannot be determined without explicit knowledge of $f(A-Z)$, but this is not so, as the following device shows. Let A_e, Z_e be a nucleus with empirically known neutron-binding energy, and with

$$N_e = A_e - Z_e = A - Z = N.$$

Then we have

$$Q_n(A_e,Z_e) = \alpha - \beta - \tfrac{2}{3}\gamma A_e^{-1/3} + 4\beta \frac{Z_e^2}{A_e^2}\left[1 + \frac{\epsilon}{12\beta} A_e^{2/3}\right] + f'(A_e - Z_e), \qquad (23a)$$

where

$$f'(A_e - Z_e) = f'(A-Z).$$

Subtracting Eq. (23a) from (23), we have

$$Q_n(A,Z) = Q_n(A_e,Z_e) + \tfrac{2}{3}\gamma(A_e^{-1/3} - A^{-1/3}) + 4\beta\frac{Z^2}{A^2}\left[1 + \frac{\epsilon}{4\beta}A^{2/3}\right] - 4\beta\frac{Z_e^2}{A_e^2}\left[1 + \frac{\epsilon}{4\beta}A_e^{2/3}\right]$$

$$\simeq Q_n(A_e,Z_e) - 4\beta\left[\frac{Z_e^2}{A_e^2} - \frac{Z^2}{A^2}\right] - 0.4. \qquad (24)$$

In the last approximate expression we have evaluated in Mev the differences in the surface and Coulomb terms for the cases which will be found to be of interest. The isotopic term in β is by far the largest term in the difference in neutron binding energy for two nuclei with the same N. The important result of the foregoing analysis is that $g(Z)$ does not appear in (23) for $Q_n(A,Z)$ and $f(N)$ and $f'(N)$ do not appear in (24). This follows from our assumptions that to first order the deviations from the Weizsäcker mass law are separable in N and Z.

If the values of $Q(A_e,Z_e)$ were always accurately known the present procedure would be simple of application. Unfortunately, individual empirical values of $Q(A_e,Z_e)$ contain sufficient inaccuracies to cause serious difficulties if employed singly. The difficulties can be overcome, in part, by a suitable form of smoothing, except for values of $A-Z$ between ~ 84 and ~ 110, in which range the empirical mass data seem to be too poorly known even to admit of smoothing.

Smoothing is effected by constructing the quantity

$$\Delta\langle f'(A_e - Z_e)\rangle_{\mathrm{Av}} \simeq \Delta\left[Q_n(A_e,Z_e) - 4\beta\frac{Z_e^2}{A_e^2}\right]_{\mathrm{Av}}.$$

The quantity in square brackets is first calculated for each nucleus for which empirical values of Q_n are available, and then averaged over all cases having the specified value of $N = A_e - Z_e$. The Δ indicates that the values are calculated relative to those at the beginning of the appropriate neutron shell. Thus $\Delta\langle f'\rangle$ is the excess neutron binding energy *to* nuclei with a specified N over that given by the smooth Weizsäcker mass formula normalized to zero at the beginning of the shell in which N lies. We have used the nuclear mass tables of Wapstra (Wa55) and Huizenga (Hu55) for this purpose. When this has been done for various N the resulting average values have been plotted against N, as in Fig. VII,1. A second form of averaging is then effected by drawing a smooth curve through the points obtained in this way, values being finally read off the curve.

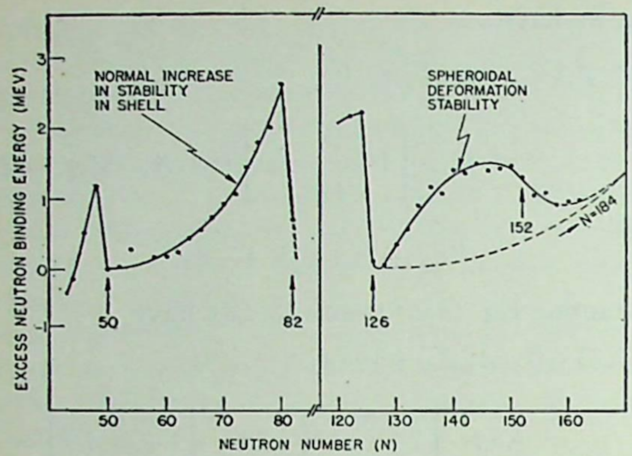


FIG. VII,1. The average excess neutron binding energy *to* nuclei with neutron number N, over that given by the smooth Weizsäcker atomic mass formula (We35). Each value has been normalized to the value at the beginning of the shell in which N lies. Empirical mass data for the $N=82$ to 126 shell are not accurate enough to yield the portion of the curve in this region.

We interpret the "quadratic" rise in the neutron binding excess between $N=50$ and 82 as the normal behavior as a neutron shell is filled. The last neutrons interact with a greater number of nucleons and are bound by about 2.6 Mev more than the first neutrons in this shell. In the $N=126$ to 184 shell there is a rapid rise to great stability early in the shell, a flattening off, and eventually a decrease having a maximum slope near $N=152$. We assume that eventually the curve rises again to a high value at shell closure. The early rise may be attributed to the ability of these heavy nuclei to take up spheroidally deformed shapes which have lower energy levels for extra neutrons than the spherical shape at the beginning of the shell. Maximum deformation is reached apparently near $A\sim 144$ and then the shell becomes spherical again on closure, with the normal shell effect coming into play. The effect of spheroidal deformations on quadrupole moments and on rotational level structure in heavy nuclei has been brilliantly developed by Bohr and Mottelson (Bo53) and extensive calculations of energy levels in spheroidal potentials have been made by Nilsson (Ni55). Nuclei in the $N=51$ to 82 shell do not show these effects in a pronounced manner, and thus "normal" shell behavior can be expected. The effects are well known in the heavier $N=82$ to 126 shell, for which, unfortunately, masses accurate enough for our purposes are not available. The effects can be expected to be highly developed in the $N=126$ to 184 shell. The influence of the enhanced stability on the r process is discussed in detail in what follows. *Note added in proof.*—The recent mass spectroscopic data of W. H. Johnson, Jr., and V. B. Bhanot (Jo57) yield a curve in the range $N=82$ to 126 similar to the right-hand curve in Fig. VII,1.

A difficulty which arises in determining the $Q_n(A,Z)$ from the empirical $Q_n(A_e,Z_e)$ lies in the isotopic parameter β for which we have taken the value 22.6 Mev. This is the value which gives the correct dependence of Z_A on A if the Coulomb coefficient ϵ is taken as 0.69 Mev. Fowler *et al.* (Fo47) found this value for ϵ to be necessary to fit the rise of the packing fraction curve for nuclei with $A>70$, and Green has confirmed this point (Gr54b). This value for ϵ is somewhat larger than that found in early attempts to adjust the smooth Weizsäcker formula to nuclear masses but is in keeping with the smaller values of nuclear radii found in electron scattering experiments by Hofstadter and his collaborators (Ho56c). However, in determining $Q_n(A,Z)$ from $Q_n(A_eZ_e)$ we really need "local" values of β at $N=A-Z=A_e-Z_e$, not the average value determined by the best fit to Z_A over all A. This implies that β is not constant, as is indeed found to be the case if one adjusts the Weizsäcker formula or its alternative form in M_A, B_A, Z_A to local regions. Unfortunately, the empirical values for $Q_n(A_e,Z_e)$ do not cover a sufficient spread in A_e and Z_e to give accurate determinations of the local values for β, and thus the extrapolation to A,Z values far from the stability line involves considerable uncertainty.

To elaborate, let us consider the accuracy with which we must compute Q_n for the waiting point, A, at a given Z value along the capture path not near any magic

TABLE VII,1. Parameters of the r process at magic neutron numbers.

N	A	Z	Process	Q_n[a] Mev	Z_A	B_A Mev	$\frac{1}{2}\delta_A$ Mev	W_β Mev	$\tau\beta$ sec	A	n(calc)	n(obs)
50	80	30	$(\beta^-\nu_-)$	1.9	35.0	1.90	1.43	3.82	12.3	1	42.6	33.8
	81	31	$(\beta^-\nu_-)$	2.4	35.4	1.88	···	4.07	8.96	3	10.4	6.62
	82	32	(n,γ)	2.9[a]				Break-through point				
82	128	46	$(\beta^-\nu_-)$	1.0	54.5	1.36	1.18	7.48	0.43	1	Stan.	1.48
	129	47	$(\beta^-\nu_-)$	1.4	54.9	1.36	···	7.84	0.34	1	1.17	1.05
	130	48	$(\beta^-\nu_-)$	1.7	55.3	1.35	1.16	5.82	1.50	3	1.73	···
	131	49	(n,γ)	2.0[a]				Break-through point				
126	194	68	$(\beta^-\nu_-)$	1.2	77.9	1.16	0.77	8.31	0.25	1	0.87	0.533
	195	69	$(\beta^-\nu_-)$	1.4	78.3	1.16	···	8.39	0.24	3	0.28	0.548
	196	70	(n,γ)	1.6[a]				Break-through point				

[a] The values for Q_n, at which (n,γ) occurs at the recognizable break-through points of the magic-neutron-number peaks in the atomic abundance curve, have been used to establish a smooth dependence of Q_n on atomic weight.

shell number in N. We evaluate

$$\left.\frac{\partial Q_n}{\partial A}\right|_Z \simeq -\frac{8\beta Z^2}{A^3} \sim -0.25 \text{ to } -0.1 \text{ Mev},$$

where the numerical values have been determined near $A=100$ and $A=200$, respectively. Since the Q_n can hardly be specified to better than 0.5 Mev an uncertainty in A of several units is to be expected. The break-through points near magic shell numbers for N can be evaluated from

$$\left.\frac{\partial Q_n}{\partial A}\right|_N \simeq +\frac{8\beta ZN}{A^3} \sim 0.5,\ 0.3,\ 0.2 \text{ Mev},$$

where the numerical values have been determined at $N=50$, 82, 126, respectively. Again an uncertainty of one or more units in A arises from uncertain knowledge of the Q_n.

With these difficulties in mind, we have chosen to establish a few key points in the track of the r process by using the values for the atomic weights at those peaks in the atomic abundance curve which are attributed to the r process. On the leading edges of these peaks there is no arbitrariness, except in the value of A (or Z) at which a given magic N is reached, and in the value of A (or Z) at which break-through occurs. The break-through points at which neutron capture without a preceding beta decay becomes possible occur just beyond the maxima of the magic-neutron-number abundance peaks. We have chosen these break-through points to give the best fit to the three abundance peaks due to magic neutron-numbers and have then adjusted the path determined by our criterion $Q_n(A,Z) \geq Q_0$ to fit the three points and to give a smooth run of waiting points between these break-through points. Thus the r-process path has been determined by a smooth curve for Q_0 passing through the break-through values at $N=50$, 82, and 126. These key values are shown in Table VII,1, where we list the Q_n values calculated by the smoothing procedures, using $\beta=22.6$ Mev. These Q_n values show a systematic decrease from 2.9 Mev at $N=50$, through 2.1 Mev at $N=82$ to 1.6 Mev at $N=126$. These values correspond to temperatures of 1.45, 1.0, and 0.8×10^9 degrees, respectively, for $n_n=10^{24}$. Whether or not this systematic decrease is real is a matter which cannot be answered with present knowledge of nuclear masses and binding energies. In attempting to use estimated local values for β we have found that the differences in Q_n and thus T are only accentuated. It is clear that the conditions, n_n and T, for the production of the three peaks were not greatly different. A small difference, if real, would be very significant in our understanding of the astrophysical circumstances under which the r-process elements were synthesized. We return to this matter later in our discussion and in Sec. XII. The capture path from Fe^{56} to

TABLE VII,2. The path of the r process and calculations of r-process abundances.

N	A	Z	Z_A	B_A Mev	$\frac{1}{2}\delta_A$ Mev	W_β Mev	τ_β sec	ΔA	Atomic abundances
30	56	26	...	...	...	...	...	...	Starting point
42	68	26	30.3	2.40	1.40	3.42	21.4	3	First waiting point
44	71	27	31.4	2.20	...	4.68	4.45	7	2.20
50	78	28	34.2	1.94	1.42	6.26	1.04	1	3.61
	79	29	34.6	1.92	...	6.45	0.90	1	3.11
	80	30	35.0	1.90	1.43	3.82	12.3	1	42.6
	81	31	35.4	1.88	...	4.07	8.96	3	10.4
52	84	32	37.4	1.80	1.44	4.28	6.96	1	24.2
	85	33	37.8	1.78	...	4.59	4.91	5	3.40
56	90	34	39.8	1.70	1.44	4.67	4.50	7	2.23
62	97	35	42.5	1.60	...	8.50	0.225	5	0.156
66	102	36	44.5	1.55	1.41	8.39	0.241	5	0.167
70	107	37	46.5	1.50	...	11.00	0.062	3	0.072
72	110	38	47.7	1.47	1.35	9.73	0.115	7	0.057
78	117	39	50.5	1.41	...	13.19	0.025	3	0.029
80	120	40	51.7	1.39	1.27	12.02	0.040	3	0.046
82	123	41	52.8	1.38	...	13.33	0.024	1	0.083
	124	42	53.1	1.37	1.23	11.05	0.061	1	0.211
	125	43	53.5	1.37	...	11.46	0.051	1	0.176
	126	44	53.8	1.37	1.20	9.30	0.144	1	0.500
	127	45	54.2	1.36	...	9.61	0.122	1	0.423
	128	46	54.5	1.36	1.18	7.48	0.427	1	1.48
	129	47	54.9	1.36	...	7.84	0.338	1	1.17
	130	48	55.3	1.35	1.16	5.82	1.50	3	1.73
84	133	49	57.6	1.33	...	8.61	0.211	3	0.244
86	136	50	57.6	1.31	1.08	6.10	1.18	9	0.454
94	145	51	60.6	1.28	...	9.59	0.123	5	0.085
126	185	59	74.9	1.18	...	16.31	0.009	1	0.030
	186	60	75.3	1.18	0.78	14.82	0.014	1	0.049
	187	61	75.6	1.18	...	14.78	0.014	1	0.049
	188	62	75.9	1.17	0.78	13.06	0.026	1	0.091
	189	63	76.3	1.17	...	13.14	0.025	1	0.088
	190	64	76.6	1.17	0.78	11.54	0.049	1	0.170
	191	65	76.9	1.16	...	11.40	0.052	1	0.180
	192	66	77.3	1.16	0.77	9.94	0.103	1	0.357
	193	67	77.6	1.16	...	9.90	0.105	1	0.364
	194	68	77.9	1.16	0.77	8.31	0.252	1	0.874
126	195	69	78.3	1.16	...	8.39	0.240	3	0.278
128	198	70	80.4	1.16	0.76	8.90	0.180	11	0.056
138	209	71	83.9	1.16	...	12.56	0.032	3	0.037
140	212	72	84.9	1.16	0.75	11.81	0.043	3	0.050
142	215	73	86.0	1.16	...	12.68	0.030	3	0.035
144	218	74	87.0	1.16	0.75	11.93	0.041	3	0.048
146	221	75	88.1	1.16	...	12.80	0.029	3	0.034
148	224	76	89.1	1.16	0.74	12.06	0.039	3	0.045
150	227	77	90.2	1.16	...	12.91	0.028	3	0.032
152	230	78	91.3	1.16	0.73	12.30	0.036	1	0.123
	231	79	91.6	1.16	...	12.22	0.037	1	0.127
	232	80	91.9	1.16	0.73	10.67	0.072	3	0.084
154	235	81	93.0	1.16	...	11.52	0.049	3	0.057
156	238	82	93.4	1.16	0.72	10.10	0.095	3	0.110
158	241	83	94.5	1.16	...	10.94	0.064	1	0.221
158	242	84	94.8	1.16	0.72	9.41	0.136	3	0.157
160	245	85	95.9	1.16	...	10.24	0.089	3	0.103
162	248	86	97.0	1.16	0.72	9.64	0.120	3	0.139
164	251	87	98.0	1.16	...	10.36	0.084	3	0.097
166	254	88	99.0	1.16	0.72	9.64	0.120	3	0.139
168	257	89	100.1	1.16	...	10.48	0.029	3	0.091
170	260	90	101.2	1.16	0.71	9.87	0.107	5	0.074[a]
174	265	91	102.9	1.16	...	11.40	0.052	7	0.026[a]

[a] Probably depleted by neutron-induced fission.

$A=265$ which we have used is given in Table VII,2 and is illustrated in Figs. V,2 and V,3. The path is complete except for a gap from $A=150$ to $A=185$, in which region the empirical masses are so poorly known that there is no guide at all even to the trend of the capture track.

B. Calculation of r-Process Abundances

We now pass to the conclusion of our problem, namely, calculation of the relative abundances of the elements produced in the r process. This involves determination of $\Delta A = \partial Z/\partial A$ and τ_β at the waiting points. The capture path calculated from the criterion $Q_n(A,Z) \geq Q_o$ yielded nonintegral values of Z and A and thus a fairly smooth $\partial Z/\partial A$ along the track which we used as a guide as discussed above. In the track we

have finally chosen in Table VII,2, and in the "actual" track, Z and A change by integral values, and so we have chosen to use the integral ΔA listed in Table VII,2, in calculating $n(A)$. In Table VII,2 the various parameters needed in the abundance calculations are also listed.

The mean beta-decay lifetimes at the waiting points have been calculated from the Fermi expression

$$\tau_\beta = \frac{1}{\lambda_\beta} = \frac{10^4}{W_\beta{}^5} \text{ sec},$$

where W_β, measured in Mev, is the effective beta-decay energy at the waiting point. It includes the kinetic energy of the electron and neutrino and the rest-mass equivalent energy (0.5 Mev) of the electron. The numerical constant has been determined by using an allowed transition $\log_{10}ft$ value of 3.85, where

$$f = \frac{1}{30}\left[\frac{W_\beta}{m_o c^2}\right]^5 \simeq W_\beta{}^5$$

if W_β is measured in Mev, and $t = 0.693\tau_\beta$ is the beta-decay half-life. The method described below for evaluating W_β leads in general to an underestimate, so that our values of τ_β could be too large by a factor of as much as 10. However, the relative values and thus the relative abundances should not be uncertain by such a large factor.

The determination of W_β as a function of the waiting point charge, Z, and mass, A, is a matter of some complexity. The total kinetic energy, Q_β, available in the beta transition to the ground state of the daughter nucleus is calculable from the atomic mass expression given in Eq. (22) and becomes

$$Q_\beta = c^2[M(A,Z,N) - M(A, Z+1, N-1)].$$

Since the masses are atomic rather than nuclear, no allowance need be made for the beta-electron mass. We thus have,

$$Q_\beta = B_A(Z_A - Z - 0.5) + \frac{dg}{dZ} - \frac{df}{dN}. \qquad (25)$$

In this expression dg/dZ and df/dN explicitly appear. However, since they appear as a difference their influence on Q_β cannot be too great. Except at shell numbers, g and f increase with Z and N, respectively. If g and f were linear in Z and N with the same coefficient, then $dg/dZ - df/dN$ would be zero. Actually, g and f are approximately quadratic in the nuclear excess over magic-number values. Since there is a neutron excess in negative beta decay, in general $df/dN > dg/dZ$ and Z_A and B_A are usually smaller than the values calculated from the smooth formula. Z_A changes discontinuously at magic numbers, increasing after N magic and decreasing at Z magic. Actually, the residual effect of these terms can be taken into account by using just the first term of Eq. (25) for Q_β with local values for B_A and Z_A, particularly for Z_A. Local values for Z_A have been given by Coryell (Co53) and we have used these values as tabulated in Table VII,2. Z_A is not only a function of A but also of the shell in which Z and N fall. Coryell chose the smooth Weizsäcker relation between local values of B_A and Z_A, namely $B_A Z_A \simeq 4\beta$, but we have not found this to be sufficiently accurate for our purpose. Accordingly, we made an independent analysis in order to obtain the values of B_A listed in Table VII,2. To do this we used Coryell's Z_A at a given A and determined B_A from the empirical Q_β differences calculable at a given A from the masses of Wapstra (Wa55) and Huizenga (Hu55). These values of B_A are shown as a function of A in Fig. VII,2.

We thus have

$$Q_\beta = B_A(Z_A - Z - 0.5),$$

where B_A and Z_A are now the "local" values for these quantities taking into account all deviations from the Weizsäcker mass law except odd-even effects, which must be treated separately, as discussed below, for nuclei with odd A and even A. It is emphasized that Z_A must be chosen to correspond to the shells in which Z and N lie.

The beta-ray transition to the ground state of the daughter nucleus will rarely be allowed. Allowed transitions occur to excited states of the daughter nucleus having spins differing by zero or unity from those of the parent ground state and having the same parity. Of these allowed transitions perhaps one is favored by a large matrix element indicating considerable similarity in the initial and final states as to spin, isobaric spin, and orbital characteristics. The excitation of this favored state will in general be lower at larger A, since the level density increases with A. It can be argued that the amount of excitation ought to be proportional to, and of the order of magnitude of, the isobaric parameter B_A. We have arbitrarily chosen $2B_A$ as the excitation energy of the excited state to which the beta transition effectively occurs. With this choice, the relative heights of the abundance peaks at $N=50$, 82, and 126 are approximately reproduced as indicated by a comparison of the columns giving n(obs) and n(calc) in Table VII,1. The choice of $2B_A$ probably errs on the high side, so that our estimates for W_β may be somewhat low.

We can now write down W_β as soon as pairing energy or odd-even terms are clarified. If δ_A is used to designate the pairing energy term, and if odd A nuclei are taken as arbitrary standards for which no pairing correction is made, then for the mass of the ground states of even A, even Z, even N nuclei we must subtract $\delta_A/2$, and for even A, odd Z, odd N nuclei we must add $\delta_A/2$. Our waiting nuclei all have even N, so that for odd A no correction is necessary and for even A we must subtract $\delta_A/2$. On the other hand, the special symmetry properties of the ground states of nuclei insofar as pairing

effects are concerned do not apply to the excited states. This has been shown by Hurwitz and Bethe (Hu51) in connection with their analysis of the neutron capture cross section of odd A and even A nuclei. Hence, we neglect pairing energy terms in the daughter nuclei. As a final result, then,

$$W_\beta = B_A(Z_A - Z - 2.5) + 0.5 \begin{array}{ll} +0 & A \text{ odd, } Z \text{ odd} \\ -\frac{1}{2}\delta_A & A \text{ even, } Z \text{ even.} \end{array}$$

This prescription leads if anything to an underestimate of W_β, as indicated above, so that our values for $\tau_\beta \sim W_\beta^{-5}$ may be high by as much as a factor of ten. The term $m_o c^2 = 0.5$ Mev has been added to give the total kinetic plus rest mass energy of the effective beta decay. For $\delta_A/2$ we use the empirical values tabulated by Coryell (Co53). His values of $\delta_A/2$ as a function of A are shown in Fig. VII,2. The resulting values of W_β are given in column 7 of Table VII,2 and have been used to calculate the values of τ_β given in column 8. These in turn have been used to calculate the abundances given in column 10, arbitrarily taking the abundance of $_{52}Te_{76}^{128}$ near the $N=82$ peak as standard at 1.48 on the Suess and Urey scale. The calculated abundances are shown as a histogram for comparison with Suess and Urey abundances which are plotted as points in Fig. VII,3. The results of the calculations, extended into the range above $A=209$, are shown in Fig. VII,4.

From Fig. VII,3 reasonable but not exact agreement with observed abundances is obtained on the basis of the theory of the r process that has been outlined. The break-through values for A were established from the rounding-off points of the three observed abundance peaks ascribed to $N=50$, 82, and 126, respectively. However, the rapid rise in abundance below the peak point is unambiguously given by the rapid decrease in beta-ray energy as the track is held in the Z, A plane to the magic N numbers until break-through is finally attained. The relative abundances of the peaks come out quite well on our calculations, but this can be attributed to our arbitrary choice of the excitation energy, $2B_A$,

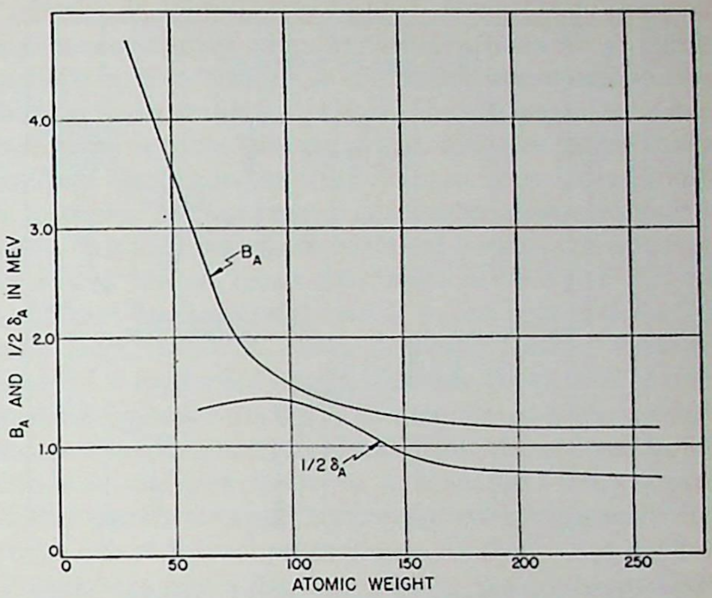


FIG. VII,2. The parameters B_A and $\delta_A/2$ as a function of atomic weight. B_A was calculated by the method described in the text, while $\delta_A/2$ is taken from the work of Coryell (Co53).

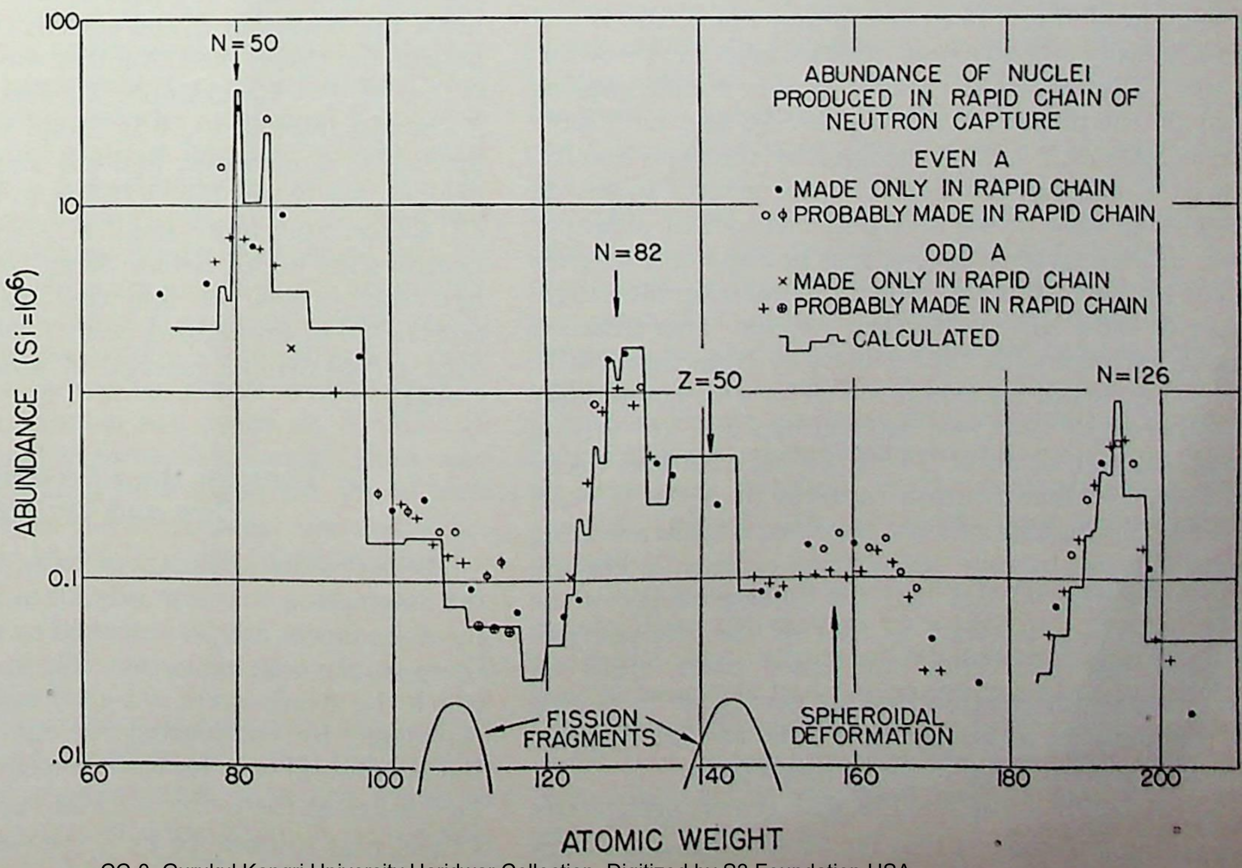


FIG. VII,3. Abundances of nuclei produced in the r process. The empirical points are taken from Suess and Urey (Su56). The histogram is a curve calculated by the methods described in the text. The free parameters have been adjusted to yield the correct relative heights of the three abundance peaks for magic neutron numbers $N=50$, 82, and 126.

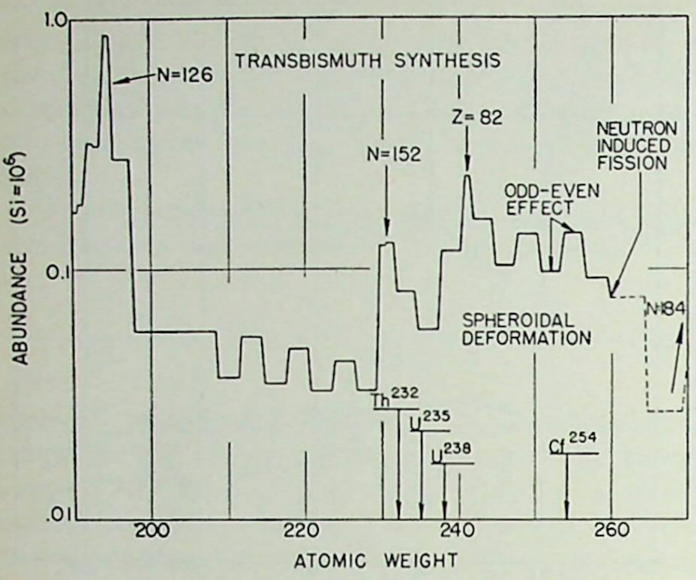


FIG. VII,4. Abundances of heavy nuclei produced in the *r* process. The method of calculation was the same as that used in calculating the abundance histogram shown in Fig. VII,3. It is assumed that neutron-induced fission terminates the *r* process at $A=260$.

in the residual nucleus after beta decay. We can conclude as follows: on the basis of the *r*-process track shown in Fig. V,2 which passes through neutron-rich nuclei with neutron-binding energies of the order of 2 Mev, and on the basis of beta-decay lifetimes for these nuclei computed from reasonable local values for the parameters B_A, Z_A, and $\delta_A/2$, it has been possible to compute a satisfactory caricature of the abundances of those nuclei which have been produced predominantly in the *r* process.

Let us discuss a few special features of the calculated histogram of Fig. VII,3. In the first place, the $N=50$ peak is really a double peak with maxima at $A=80$ and $A=84$. The empirical validity of this double peaking rests on the ratios $Se^{80}/Se^{82}=33.8/5.98$ and $Kr^{83}/Kr^{84}/Kr^{86}=5.89/29.3/8.94$, determined by the mass spectroscope. Changing the Se/Kr ratio, which might be possible in view of the divergent methods of obtaining their individual abundances, does not lead to a solution of the problem. It is also difficult to question the assignment of these isotopes to the *r* process. This situation can be caricatured in the *r* process by (i) letting $A=80$, $Z=30$ be the last nucleus at $N=50$ for which beta decay is followed by only one neutron capture; (ii) letting break-through occur at $A=81$, $Z=31$, with beta decay followed by three neutron captures to $A=84$, $Z=32$, $N=52$; (iii) making this last described nucleus a waiting point followed by only one neutron capture, in keeping with the empirical fact that the binding energy of the 53rd neutron is as small as that of the 51st neutron. Beyond this point the abundance curve trails off gradually as the neutron binding energy for increasing N gradually increases up to $N=82$ and the capture process slowly returns to the off-peak path.

The backside of the $N=126$ peak is very precipitous, in keeping with the fact that the neutron-binding energy rapidly increases after the drop at $N=126$, and the rapid return to the very neutron-rich nuclei gives a small dZ/dA ($dN/dA\sim1$), a small τ_β, and hence a small abundance. As noted previously, the rapid increase in neutron-binding energy after $N=126$ can be attributed to the fact that the first neutrons in the $N=127$ to $N=184$ shell can distort the closed nucleon shells ($N=126$, $Z=82$) in such a way as to lead to a spheroidal deformation considerably more stable than the spherical shape at the end of the preceding shell. When maximum deformation is reached, further neutron addition does not lead to greater stability, and the neutron-binding energy curve reaches a plateau. This results in the *r*-process path moving somewhat closer to the stability line in the Z, A plane. This leads in turn to somewhat greater values of τ_β and dZ/dA and thus to a moderate increase in abundance. It is to this spheroidal deformation effect that we attribute the abundance hump from $A=230$ to 260 in Fig. VII,4. Additional discussion of this feature is given in Sec. VIII. We also attribute the abundance hump near $A=160$, which is evident in Fig. VII,3, to this same effect, even though we have not been able to make detailed calculations in this region. It is well known that the stable nuclei in this region show large quadrupole moments consistent with the fact that the most stable forms of the ground states are spheroidal rather than spherical in shape.

The precipitous drop expected on the backside of the $N=82$ peak is masked by the appearance of a peak due to the magic number of protons, $Z=50$. It might be thought that in a neutron capture process, proton magic numbers would be irrelevant. However, a proton magic number does affect the beta-decay lifetime by slowing down the transition to the nucleus containing a proton outside the closed shell. Another way of expressing this is to note that, for example, Z_A does not increase when Z changes from 49 to 50 in Table VII,2, and thus W_β decreases, τ_β increases, and the abundance hump evident at $A=140$ in Fig. VII,3 is to be expected. Since dZ/dA becomes quite small at Z magic, as numerous neutrons are added before decay to the proton in the next shell occurs, we find that the abundance value established at the waiting point ($A=136$) is spread out over a considerable interval of atomic weights up to $A=145$.

C. Time for the *r* Process: Steady Flow and Cycling

The *r*-process abundances have been calculated on the assumption that a steady state is reached in which the abundances are proportional to the beta-decay lifetimes at the waiting points. The circumstances under which the steady state is a good approximation can be elucidated by considering the individual beta decay times and total time for the *r* process. From Table VII,2 Fe^{56}, taken as the source nucleus, rapidly captures some 12 neutrons to produce Fe^{68}, the first waiting point in

the "typical" capture path. Here Q_n is less than Q_o ($\sim$3 Mev in this region) for the first time and beta decay occurs with a lifetime of $\sim$21 sec. This is longer than all the succeeding lifetimes and is in fact $\sim$25% of the total time for the *r* process. This latter quantity is just the sum of all the τ_β in Table VII,2, plus a rough estimate of 3 sec for the region $A=150$ to 185, and is therefore $\sim$80 sec. After passing Fe^{68}, the material spends $\sim$40 sec in the $N=50$ (and 52) peak. Thus the long-lived "source" and first "sink" control the flow for a considerable period during which the heavier, shorter-lived products come into, and remain in equilibrium with, these long-lived predecessors. It is significant that the Fe^{68} lifetime plus the $N=50$ peak lifetime is long as compared with the 5 to 23 sec lifetimes of Mg^{22}, Na^{21}, etc., which control the production of the neutrons. It is also significant that the total time for the *r* process is comparable to the $\sim$100-sec expansion and cooling time for the supernova envelope.

From the correspondence of the calculated abundances with those in the atomic abundance distribution, it would seem that this distribution was produced under conditions in which steady flow was just reached and maintained. Other conditions can pertain in different circumstances. The establishment of steady flow requires about 38 neutrons per capturing Fe^{56}, since the average atomic weight in the distribution given in Fig. VII,3 is $\simeq$94. For fewer than 38 neutrons per Fe^{56}, the upper end of the abundance curve will not reach its full value. For a greater number of neutrons, as will certainly be the case in some supernovae, the abundances of high atomic weight will be enhanced over those illustrated in Fig. VII,3. This enhancement occurs over the entire region above $A\sim110$ because of cycling resulting from the neutron-induced fission process near $A\sim260$. For asymmetric fission, one-half of the flow is returned near $A\sim110$ and the full flow is maintained above $A\sim150$. Thus the abundances above $A\sim110$ grow relative to those below $\sim$110, with a factor of two marking the growth above $A\sim150$ relative to that for $A\sim110$ to 150. For a neutron to iron-group abundance ratio $\gtrsim100$, practically all of the original "seed" nuclei will be cycling in the region $A\sim110$ to 260 and increasing in abundance with average atomic weight $A\simeq174$. Under these circumstances the peaks at $N=82$ and $N=126$ have about the same abundances, since only one half of the material is returned to the $N=82$ peak in the cycling. In Sec. XII we consider the case of enhanced production above $A\sim110$ to be typical for certain supernovae.

D. Freezing of the *r*-Process Abundances

Implicit in our comparison of calculated abundances with those observed is the assumption that the unstable nuclei produced at mass number A in the *r* process subsequently decay only by beta decay to their stable isobars with the same mass number. This is indeed the predominant final step in the synthesis of a stable nucleus by the *r* process as the temperature and neutron flux decrease and the *r* process abundances are "frozen in." In some cases beta decay can occur to a state of the residual nucleus in the final chain of beta decays which is so highly excited that it is unstable to neutron emission. This is similar to the delayed emission of neutrons following fission. Since this mode of decay is not the main one, we have neglected it in our calculations. We note here, however, that delayed neutron emission occurs for even-even nuclei, since the depression of their ground states results in the availability of higher excited states when they are produced. Thus, this process favors odd A production over even A, since some even A mass chains switch to odd by delayed neutron emission. This point tends to compensate for the fact that even A abundances are somewhat greater than odd A abundances before the freezing occurs.

Smart (Sm48, Sm49) suggested, in connection with primordial synthesis of the elements, that (γ,n) reactions compete with negative beta decay in the freezing after a rapid neutron capture process. In this way the shielded isobars could be produced. In stellar synthesis the shielded isobars are produced separately in a second process which we have labeled the *s* process and which we suggest has occurred during the giant stage of stellar evolution. It is our belief that (γ,n) reactions subsequent to the *r* process, primordially or in supernovae, cannot have produced the shielded isobar abundances found in the atomic abundance curve.

It has been noted in Sec. VI that the barium isotope ratios and absolute abundances are consistent with *p*- and *s*-process production. Let us now consider the difficulties arising in an attempt to explain the observed barium abundances, for example, on the basis of (γ,n) reactions in the freezing of the *r* process. In the first place the possibility can be considered that the neutron flux suddenly ceases while the temperature is still high and that (γ,n) reactions rather than negative beta decays return the material to the stability line. This would require the ejection of some 10 to 30 neutrons per product nucleus and would obviously be self-defeating in that it would almost restore the original neutron flux. Furthermore it would invalidate the explanation for the *r*-process peak abundances given in this section and would require double abundance humps to have been produced in the *r*-process path near $A\sim100$, 140, and 200, for which there are no known reasons.

A more reasonable possibility is that suggested by Smart; after the original very fast beta decays bring the nuclei close to the stability line, the beta lifetimes become long and the (γ,n) process competes in the final adjustments. Thus in the case of barium, material produced at A just greater than 138 might stop, after fast beta decay, in the long-lived unstable barium isotopes ${}_{56}Ba_{83}{}^{139}$ (85 min), ${}_{56}Ba_{84}{}^{140}$ (12.8 days), and ${}_{56}Ba_{85}{}^{141}$ (18 min). Then (γ,n) processes would scour out the

loosely bound 85th, 84th, and 83rd neutrons and pile up the material in ${}_{56}Ba_{82}{}^{138}$ with its tightly bound 82nd neutron. Some leakage to the lighter isotopes could occur and a general tailing off in abundance from ${}_{56}Ba_{82}{}^{138}$ to ${}_{56}Ba_{74}{}^{130}$ would be expected. This is indeed found with the important exception that ${}_{56}Ba_{74}{}^{130}$ is even slightly more abundant than ${}_{56}Ba_{76}{}^{132}$, which is not to be expected at all. Furthermore, if the (γ,n) process does occur at barium, why does it not occur for xenon? On the basis of this process one would expect magic ${}_{54}Xe_{82}{}^{136}$ to be the most abundant of the xenon isotopes, with a tailing off to ${}_{54}Xe_{70}{}^{124}$. This is not found to be the case. ${}_{54}Xe_{78}{}^{132}$ is the most abundant isotope of xenon and is 2 to 3 times as abundant as ${}_{54}Xe_{80}{}^{134}$ and ${}_{54}Xe_{82}{}^{136}$. On the other hand, this is just what is to be expected if it is supposed that the *s* and *r* processes are operating separately. ${}_{54}Xe_{78}{}^{132}$ is the heaviest stable isotope of xenon which can be produced in the *s* process, and presumably it has the smallest (n,γ) cross section. ${}_{54}Xe_{80}{}^{134}$ and ${}_{54}Xe_{82}{}^{136}$ are made only in the *r* process, and they are low in abundance in keeping with the general ratio of the *r*-process abundances to the *s*-process abundances in this region.

Returning to the case of barium, we must finally emphasize that in the Suess and Urey distribution Ba^{138} is very abundant compared to its neighbors. If it were made by (γ,n) reactions following the *r* process, this would require an abundance peak in the primary path at A just greater than 138, for which no explanation can be given in terms of known "magic" properties. On the other hand its great production in the *s* process is understood in terms of its own magic properties.

VIII. EXTENSION AND TERMINATION OF THE *r* PROCESS AND *s* PROCESS

A. Synthesis of the Naturally Radioactive Elements

Because of its long time-scale and because it builds through the beta-stable nuclei, the *s* process is incapable of producing the naturally radioactive elements beyond bismuth. Bismuth $(Z=83)$ is the heaviest of the elements with an isotope which does not decay by electron emission, alpha-particle emission, or fission. As is discussed in detail later, the short-lived alpha-emitting isotopes of polonium, which are produced after bismuth, terminate the *s* process by cycling its products back to lead.

On the other hand, just because of its short time-scale and because it builds through charge-poor, neutron-rich nuclei, the *r* process can produce transbismuth elements. These proton-poor nuclei are relatively more stable to alpha emission and fission than their stable isobars at a given A.

A convincing experimental demonstration of this was given by the production of Cf^{254} and other nuclei in the November, 1952, thermonuclear hydrogen bomb test at Bikini, as reported by Fields *et al.* (Fi56). These nuclei were synthesized by the instantaneous irradiation of U^{238} by an intense neutron flux. The successive addition of neutrons to U^{238} built U^{254} and other neutron-rich uranium isotopes. After the explosion, the U^{254} decayed rapidly to beta-stable Cf^{254}. The Cf^{254} was detected in the test debris through its spontaneous decay by the fission process. The energy release in the spontaneous fission is 220 Mev per decay, which is very large compared to the energy emitted in alpha decay (~5 Mev) or beta decay (~1 Mev). Fields *et al.* reported the half-life of Cf^{254} as 55 days. More recent measurements by Thompson and Ghiorso (Th57) yield a preliminary value of 61 days while Huizenga and Diamond (Hu57) have recently obtained a value of 56.2 ± 0.7 days.§ The Cf^{254} is unique in that it is the only beta-stable nucleus which decays predominantly by spontaneous fission with a half-life between a few days and 10^4 years.

The authors and their collaborators (Bu56, Ba56) have associated the *r* process astrophysically with the explosion of supernovae of Type I. A characteristic feature of supernovae of Type I is that after an initial period of 50–100 days the light curve exhibits an exponential decline corresponding to about 0.0137 magnitudes daily or a half-life of 55 days. The uncertainties in this half-life are discussed in Sec. XII. The close correspondence between this value and that for Cf^{254} suggested that Cf^{254} was produced in the supernova explosion along with the other products of the *r* process. Because of its large energy release per decay it was supposed in this earlier work that the Cf^{254} would dominate the radioactive energy input into the supernova debris for the order of a year after the original outburst. In the following part of this section we discuss the radioactive decay of the *r*-process products which give the radioactive energy input for Type I supernovae. The connection between the energy input and the light curve for Type I supernovae is discussed in Sec. XII. The source of the great flux of neutrons required in supernovae has been discussed in Sec. III.

B. Extension and Termination of the *r* Process

(1) *Radioactive Energy Input in Type I Supernovae*

Certain of the calculations necessary for computing the radioactive energy release by *r*-process products have been made by Schuman (Sc57a). He has analyzed the mode of decay and calculated the total decay energy for the longest-lived neutron-rich nuclei produced in each mass chain where the *r* process extends beyond the stable nuclei $(A>209)$. His results, modified by new data, are incorporated in Table VIII.1. The total decay

§ The calculations which are described in this section have all been made using a half-life of Cf^{254} of 61 days, since the authors were not aware of the work of Huizenga and Diamond when they were made. Clearly, the value obtained by them is in better agreement with the original result of Baade for the half-life of the light curve of the supernova in IC 4182.

energy includes all subsequent shorter-lived activities which quickly come into equilibrium with the decay listed. Neutrino kinetic energy is not included; 45% of the energy released in each beta decay is assigned to the electrons. In addition we list in Table VIII,1 all mass chains produced in the r process below $A=209$ which beta decay with maximum half-lives in excess of 3 days. In the table all chains with maximum half-lives less than 3 days are neglected, except for the fast spontaneous fission decays at the end of the table, which are designated (SF) in column 4; (n,f) indicates neutron-induced fission.

Table VIII,1 illustrates the fact that after the rapid beta decays in a given mass chain which immediately follow the termination of an r-process event, the beta-stable nuclei above bismuth decay either by fission to the middle of the periodic table, or by alpha emission and further beta decay to the long-lived nuclei listed. Those above uranium then decay to Th^{232} ($4n$ naturally radioactive series), Np^{237} ($4n+1$ series), U^{238} ($4n+2$ series), and U^{235} ($4n+3$ series), and eventually to Pb^{208}, Bi^{209}, Pb^{206}, and Pb^{207}. *The tabulated results confirm the uniqueness of* Cf^{254}. *It is the only beta-stable, neutron-rich nucleus which decays predominantly by spontaneous fission with a half-life between 10 days and* 10^4 *years.* The table clearly indicates that the energy released by any fission decay exceeds the energy release of any alpha-plus-beta chain by a factor of five, and thus qualitatively the dominance of the radioactive energy release by Cf^{254} is assured.

It is important to consider the empirical evidence which underlies this matter of the uniqueness of Cf^{254}. Ghiorso (Gh55) has presented a graphic display of this evidence by plotting spontaneous fission and alpha-emission half-lives *versus* neutron numbers for the transuranium nuclei. Ghiorso's diagram is shown in Fig. VIII,1. The spontaneous fission half-lives decrease precipitously in value for all elements beyond the "magic" neutron number $N=152$. In what way 152 is magic, we discuss below. For $N<152$, the fission half-lives for a given Z show the characteristic rise with increasing N, to be expected from the well-understood increase in lifetime with decreasing value of the fission parameter, $Z^2/A=Z^2/(Z+N)$ (Bo39). This parameter is proportional to the ratio of the nuclear Coulomb energy to the nuclear surface energy. The alpha-emission half-lives also decrease at $N=152$, but then recover and resume their normal increase with N, since the essential nuclear bindings increase with N, while the Coulomb energy, upon which the alpha-particle decay energy is linearly dependent, remains essentially constant. The result is the sudden occurrence, within one isotope or two for each element, of spontaneous fission as the predominant mode of decay rather than alpha emission. Ghiorso's diagram, with its ordinate of some 28 factors of ten in half-life, shows that the sudden dominance of this mode of decay occurs over a wide spectrum of half-lives for the various transuranium elements.

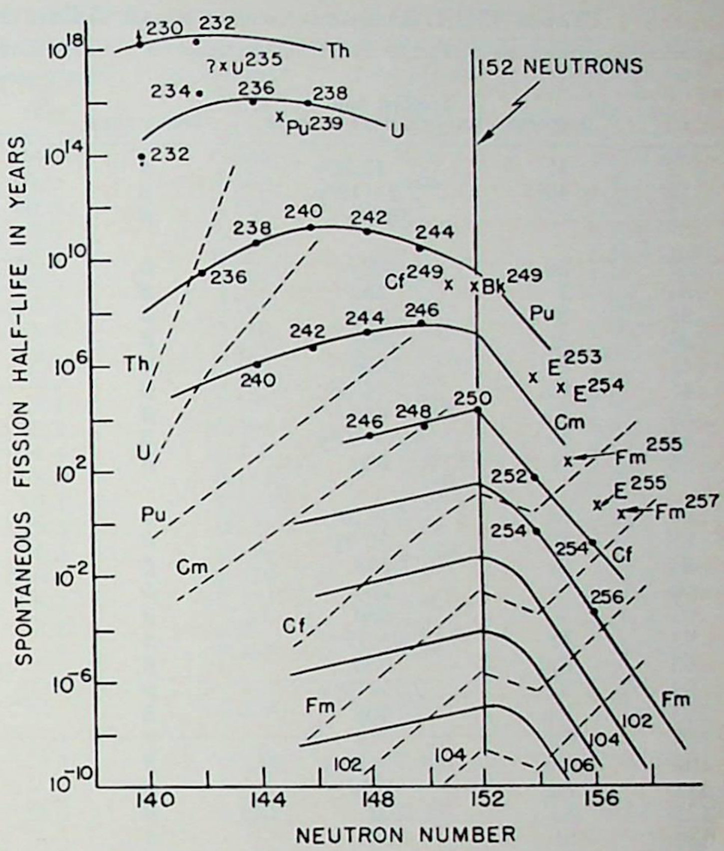


FIG. VIII,1. The analysis by Ghiorso (Gh55) of the spontaneous fission half-life (solid lines) against neutron number. Dotted lines indicate the experimentally observed alpha half-life variation except in the cases of elements 102 and 104.

Only for $_{98}Cf^{254}$ does it occur when the fission lifetime is anywhere near 55 days. For fermium it occurs at Fm^{256} with a fission lifetime of 3 hours (measured), while for curium it occurs at $_{96}Cm^{250}$ with a fission lifetime of 10^4 years (estimated). For the other relevant californium isotopes, Cf^{250} has $t_\alpha=3\times10^4$ years and $t_{SF}=10^4$ years, so that it decays by spontaneous fission a considerable proportion of the time, while Cf^{252} has $t_\alpha=2.2$ years and $t_{SF}=66$ years, so that it decays 3% of the time by spontaneous fission. For Cf^{254}, on the other hand, $t_{SF}=56$ days while $t_\alpha\geq100$ years (estimated). The fission activity predominates to the extent that alpha decay has not yet been observed for Cf^{254}.

As an empirical rule the spontaneous fission lifetime for $N\geq152$ is a function primarily of A. Thus it is 2×10^9 sec for $A=252$, 5×10^6 sec for $A=254$, 10^4 sec for $A=256$, ~10 sec for $A=258$, and $\sim10^{-3}$ sec for $A=260$.

To obtain a quantitative measure of the energy released by radioactivity by the products of the r-process event, we must multiply Schuman's calculations and those which we have made of the energy release for each mass chain, by an estimate of the relative production of the fast-decaying progenitor of that chain in the event itself. This is done in Table VIII,1, where the last columns give the total energy released, and the initial energy released measured in Mev/day. In his original calculations Schuman assumed that equal amounts of

TABLE VIII,1. Radioactive energy release of mass chains produced in the r process with half-life >3 days.

A	Element	Longest half-life, seconds, days or years	Decay type	Decay energy Q Mev	Initial abund. n_0 (Si $=10^6$)	Total energy n_0Q Mev	Initial en. rate $0.7\ n_0Q/t_{1/2}$ Mev/day
3	H	12.26y	β	0.01		for reference only	
10	Be	2.7×10^{6}y	β	0.25		for reference only	
14	C	5600y	β	0.08		for reference only	
32	Si	$\sim$300y	β	0.82	50	40.8	$\sim 1.1\times10^{-4}$
33	P	25d	β	0.12	50	5.95	0.165
35	S	87d	β	0.08	50	3.75	0.0299
39	A	260y	β	0.22	50	12.8	9.3×10^{-5}
45	Ca	160d	β	$\sim$0.11	1	$\sim$0.11	$\sim 4.8\times10^{-4}$
47	Ca	4.7d	β	1.83	100	183	26.9
59	Fe	45d	β	1.36	30	40.8	0.628
60	Fe	$\sim 3\times10^{5}$y	β	$\sim$0.05	30	$\sim$0.15	$\sim 9.4\times10^{-9}$
63	Ni	80y	β	$\sim$0.03	10	$\sim$0.3	$\sim 7\times10^{-6}$
79	Se	7×10^{4}y	β	$\sim$0.08	3.1	$\sim$0.248	$\sim 6.7\times10^{-9}$
85	Kr	10.4y	β	0.31	3.4	1.05	1.9×10^{-4}
89	Sr	54d	β	0.66	3.4	2.24	0.0287
90	Sr	28y	β	1.26	2.2	2.77	1.9×10^{-4}
91	Y	58d	β	0.70	2.2	1.54	0.0183
93	Zr	9×10^{5}y	β	$\sim$0.03	2.23	$\sim$0.067	$\sim 1.4\times10^{-10}$
95	Zr	65d	β	1.73	2.2	3.81	0.0408
99	Tc	2.1×10^{5}y	β	$\sim$0.13	0.156	$\sim$0.02	$\sim 1.8\times10^{-10}$
103	Ru	40d	β	0.6	0.167	0.10	0.0017
106	Ru	1y	β	1.61	0.17	0.274	5.2×10^{-4}
107	Pd	7×10^{6}y	β	$\sim$0.02	0.072	$\sim$0.0014	$\sim 4\times10^{-13}$
111	Ag	7.5d	β	$\sim$0.5	0.57	$\sim$0.28	$\sim$0.026
125	Sn	10d	β	$\sim$1.2	0.176	$\sim$0.21	$\sim$0.015
126	Sb	28d	β	$\sim$0.9	0.50	$\sim$0.45	$\sim$0.011
127	Sb	3.88d	β	$\sim$0.8	0.42	$\sim$0.34	$\sim$0.06
131	I	8.05d	β	0.50	1.73	0.865	0.0745
132	Te	3.21d	β	2.5	1.73	4.32	0.933
133	Xe	5.27d	β	$\sim$0.2	0.24	$\sim$0.048	$\sim$0.0063
135	Cs	2×10^{6}y	β	$\sim$0.1	0.244	$\sim$0.024	$\sim 2.3\times10^{-11}$
137	Cs	30y	β	$\sim$0.55	0.45	$\sim$0.25	$\sim 1.6\times10^{-5}$
140	Ba	12.8d	β	$\sim$2.4	0.45	$\sim$1.08	$\sim$0.058
141	Ce	32d	β	$\sim$0.27	0.45	$\sim$0.12	$\sim$0.0026
143	Pr	13.8d	β	$\sim$0.42	0.45	$\sim$0.189	$\sim$0.0095
144	Ce	285d	β	$\sim$1.5	0.45	$\sim$0.68	$\sim$0.0016
147	Nd	11.6d	β	0.45	0.09	0.040	0.0024
151	Sm	80y	β	$\sim$0.045	0.09	$\sim$0.004	$\sim 1\times10^{-7}$
155	Eu	1.7y	β	$\sim$0.12	0.1	$\sim$0.012	$\sim 1.3\times10^{-5}$
156	Eu	15d	β	$\sim$1.2	0.14	$\sim$0.17	$\sim$0.0078
161	Tb	7d	β	$\sim$0.3	0.15	$\sim$0.045	$\sim$0.0045
166	Dy	3.42d	β	$\sim$1	0.11	$\sim$0.11	$\sim$0.0223
169	Er	9.4d	β	$\sim$0.15	0.1	$\sim$0.015	$\sim$0.0011
171	Tm	1.9y	β	$\sim$0.05	0.05	$\sim$0.002	$\sim 2\times10^{-6}$
175	Yb	4.2d	β	$\sim$0.23	0.03	$\sim$0.007	$\sim$0.0011
177	Lu	6.8d	β	$\sim$0.2	0.03	$\sim$0.006	$\sim 6.1\times10^{-4}$
181	Hf	46d	β	$\sim$0.5	0.03	$\sim$0.015	$\sim 2.3\times10^{-4}$
182	Ta	112d	β	0.8	0.057	0.046	2.85×10^{-4}
183	Ta	5.2d	β	$\sim$0.6	0.03	$\sim$0.018	$\sim$0.0024
185	W	74d	β	0.19	0.031	0.006	6.0×10^{-5}
188	W	65d	β	1.0	0.09	0.09	9.6×10^{-4}
189	Re	$\sim$200d	β	0.1	0.09	0.009	$\sim 3.1\times10^{-5}$
191	Os	16d	β	$\sim$0.15	0.18	$\sim$0.027	$\sim$0.0012
194	Os	$\sim$2y	β	1.01	0.88	0.889	$\sim 8.4\times10^{-4}$
196	Ir	9.7d	β	$\sim$0.6	0.278	$\sim$0.17	$\sim$0.012
199	Au	3.15d	β	$\sim$0.23	0.056	$\sim$0.013	$\sim$0.0028
203	Hg	48d	β	$\sim$0.23	0.056	$\sim$0.013	$\sim 1.9\times10^{-4}$
210	Pb	20y	α+β	5.8	0.17	0.986	9.36×10^{-5}
222	Em	3.8d	α+β	23.1	0.034	0.785	0.143
223	Ra	11.6d	α+β	28.6	0.034	0.972	0.058
224	Ra	3.6d	α+β	28.7	0.045	1.29	0.248
225	Ra	14.8d	α+β	27.9	0.045	1.26	0.059
226	Ra	1620y	α+β	31.6	0.045	1.42	1.66×10^{-6}
227	Ac	22y	α+β	34.8	0.032	1.11	9.57×10^{-5}
228	Ra	6.7y	α+β	35.2	0.032	1.13	$3.2\ \times10^{-4}$
229	Th	7300y	α+β	33.8	0.032	1.08	$2.8\ \times10^{-5}$
230	Th	8×10^{4}y	α+β	36.4	0.123	4.48	1.06×10^{-7}
231	Pa	$3.4\ \times10^{4}$y	α+β	39.9	0.127	5.07	2.83×10^{-7}
232	Th	1.39×10^{10}y	α+β	39.3	0.641	25.2	3.43×10^{-12}

TABLE VIII,1.—*Continued.*

A	Element	Longest half-life, seconds, days or years	Decay type	Decay energy Q Mev	Initial abund. n_0 (Si $=10^6$)	Total energy n_0Q Mev	Initial en. rate $0.7\ n_0Q/t_{\frac{1}{2}}$ Mev/day
233	U	1.62×10^5y	$\alpha+\beta$	38.9	0.084	3.27	3.83×10^{-8}
234	U	2.50×10^5y	$\alpha+\beta$	41.2	0.084	3.46	2.63×10^{-8}
235	U	7.1×10^8y	$\alpha+\beta$	44.6	0.663	29.6	7.93×10^{-11}
236	U	2.39×10^7y	α	4.6	0.167	0.768	6.10×10^{-11}
237	Np	2.2×10^6y	$\alpha+\beta$	43.9	0.617	27.1	2.34×10^{-8}
238	U	4.51×10^9y	$\alpha+\beta$	46.3	0.405	18.8	7.90×10^{-12}
239	Pu	2.43×10^4y	α	5.2	0.267	1.39	1.08×10^{-7}
240	Pu	6.6×10^3y	α	5.3	0.110	0.583	1.68×10^{-7}
241	Am	$470y$	α	5.6	0.221	1.24	5.01×10^{-6}
242	Pu	3.8×10^5y	α	5.0	0.260	1.30	6.49×10^{-9}
243	Am	$7600y$	$\alpha+\beta$	5.4	0.157	0.848	2.0×10^{-7}
244	Pu	7.5×10^7y	$\alpha+\beta$	15.2	0.393	5.97	1.41×10^{-10}
245	Cm	1.1×10^4y	$\alpha+\beta$	11.4	0.430	4.90	8.43×10^{-7}
246	Cm	$4000y$	α	5.5	0.103	0.566	2.69×10^{-7}
247	Cm	$>4\times10^7y$	$\alpha+\beta$	16.3	0.103	1.68	$<7.76\times10^{-10}$
248	Cm	4.3×10^5y	α, SF(10%)	23.4	0.236	5.52	2.62×10^{-8}
249	Cf	$470y$	α	6.3	0.139	0.876	3.45×10^{-6}
250	Cm	7.5×10^3y	α, SF(75%)	170	0.139	23.6	5.97×10^{-6}
251	Cf	$700y$	α	6.2	0.236	1.46	3.96×10^{-6}
252	Cf	$2.2y$	α, SF(3%)	12.9	0.097	1.25	0.0011
253	E	$20d$	$\alpha+\beta$	6.7	0.188	1.26	0.0436
254	Cf	61 or $56.2d$	SF	220	0.139	30.6	0.349
255	E	$24d$	$\alpha+\beta$	7.3	0.139	1.01	0.0292
256	All	$\sim10^4s$	SF	...	...	...	...
257	Fm	$\sim10d$	α, SF(6%)	18.5	0.091	1.68	~0.116
258	All	$\sim10s$	SF, (n,f)	...	...	...	...
259	All	$\sim10s$	SF, (n,f)	...	...	...	...
260	All	$\sim10^{-2}s$	SF, (n,f)	...	...	...	...

odd A nuclei were produced, and that equal amounts of even A nuclei were produced, and that the ratio of production of odd to even A was 1:3. We feel that this assumption can be improved upon by using the results of Table VII,2. Table VII,2 includes the abundances which we have calculated for the r process in the range $A=71$ to 209. Estimates to be discussed later have been made for $A\leq70$. Above $A=209$ we include our calculations for the production of transbismuth mass numbers using exactly the same methods as those which were employed to calculate the abundance histogram shown in Fig. VII,3. The extended results are shown graphically in Fig. VII,4.

Some comments on the calculations above $A=209$ are appropriate. No guiding landmarks such as the observed magic number peaks are available above $A=209$. On the other hand, the relative masses of the nuclei near the beta-stable region are quite well established through extensive measurements of the alpha- and beta-decay energies of the naturally occurring and artificially produced radioactive nuclei from $A=210$ to ~260. This is amply demonstrated by the smoothness of the neutron binding energy excess curve shown in Fig. VII,1 for neutron numbers above $N=126$. The general increase in the neutron binding energy following the closure of the $N=126$ shell is apparent, as well as the rapid and considerable enhancement of stability which is to be attributed to the ability of the nucleons in nuclei with $N>126$ to take up spheroidal as contrasted to spherical distributions. The loss in neutron stability, after the maximum spheroidal deformation is reached at $N\sim140$ to 150, leads to a pseudo-shell effect‖ near $N=152$ for which there is considerable experimental evidence (Fig. VIII,1) in the sudden onset of spontaneous fission itself at this neutron number. The decrease in neutron binding energy leads to increased nuclear masses, and the energy available for fission suddenly increases. This increase, coupled with the already deformed shape which is the first step toward fission, leads to the catastrophic decline in spontaneous fission lifetimes.

Referring to Fig. VIII,3, the calculated and observed abundance curves just above the $N=126$ shell peak at $A=194$ show the precipitous drop expected from the rapid return to stability arising from the fact that the first neutrons in the $N=126-184$ shell deform these nuclei spheroidally. The greater neutron binding energy in the beta-stable region means that neutrons are readily captured, and that the capture path at $Q_n\sim2$ Mev passes through nuclei far removed from the stability line, the beta-decay energy at waiting points is enhanced, and the lifetimes and the resulting abundances are decreased ($A=200$ to $A=230$). Eventually, however, the maximum deformation is passed, and at $N=152$ the neutron binding energy is decreasing rapidly

‖ The next magic number expected after 126 on the shell model with strong spin-orbit splitting is $N=\frac{1}{3}[(n+1)^3+5(n+1)]=184$ for the $n=7$ shell.

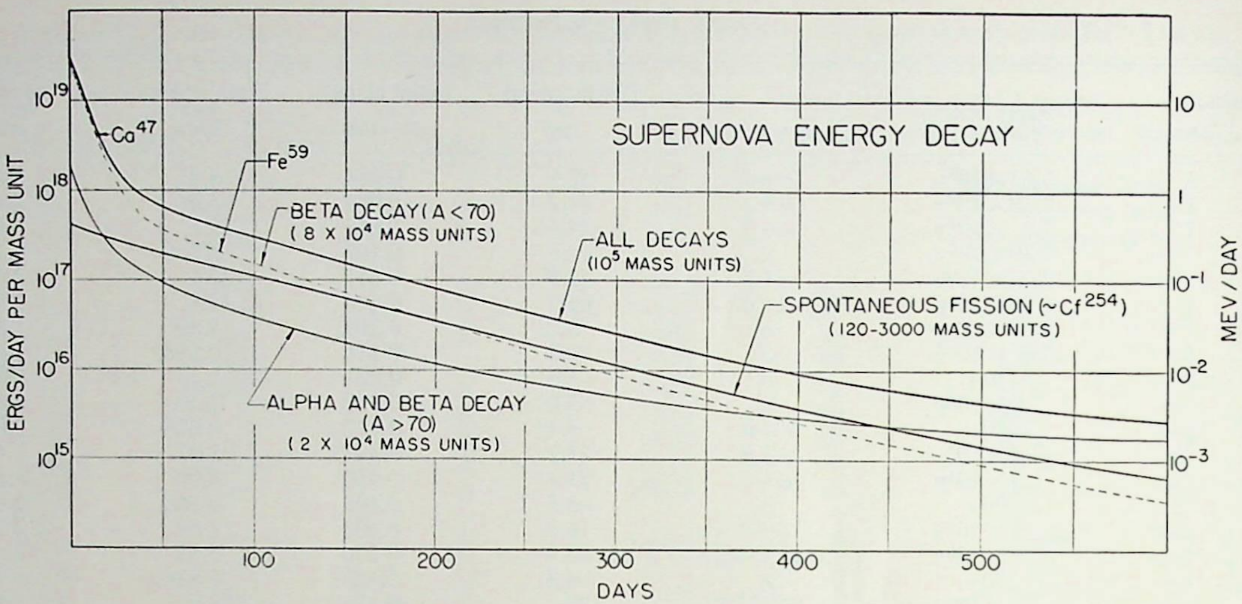


FIG. VIII,2. The radioactive energy release of the products of the r process. Radioactivity with half-lives less than 3 days has been neglected. Individual curves are shown for the beta decays of products with $A<70$, for alpha and beta decay for the products with $A>70$, for the products which decay by spontaneous fission, for the sum of all of these cases, and for Fe^{59} alone. The label for each curve includes the total mass in atomic mass units which produces the energy release measured in Mev/day on the right-hand ordinate. The scale of energy release in ergs/day, given in the left-hand ordinate, is for an amount of material measured in grams and equal numerically to the number of mass units indicated.

(Fig. VII,1), the r-process path moves toward the stability line, the beta decay waiting energy decreases, and the broad hump in the abundance curve from $A=230$ to ~ 260 corresponding to $N=152$ to 170 occurs. Included in this abundance hump is the progenitor of Cf^{254} at $A=254$. Superimposed on this broad hump is the minor peak expected at the closure of the $Z=82$ shell. Our calculations do not take into account the influence of the closing of the proton shell at $Z=82$ upon the neutron spatial distribution, which may tend to become more spherical at this point. This effect may well enhance, for the neutron-rich nuclei, the pseudo-shell effects at $N=152$, since some nuclei which lie in the r-process path fall close to having $N=152$ and $Z=82$ simultaneously.

Above $N=170$, $A=260$ the closing of the shell as N advances toward 184 would be expected to yield a deep minimum in the abundance curve just before a rise to another major peak at $N=184$, $A\sim 280$. However, the r process terminates before this minimum and subsequent rise can occur. This is because of the onset of neutron-induced fission (n,f) near $A\sim 260$, which will become more probable than neutron capture (n,γ) and will shunt the r-process products back to the middle of the periodic table at $A\sim 110$ and ~ 150, if we assume that asymmetric fission is the rule. These fission fragments are swept up in the general flow of the r process and in this way a cycling, steady flow will be established.

The sudden onset of neutron-induced fission is to be expected for exactly the same reasons as those previously enunciated to explain the onset of spontaneous fission. Even if neutron-induced fission were ignored, the 10^{-3} sec spontaneous fission lifetime will destroy the $A=260$ progenitors in an explosion for which the time-scale is 10–100 seconds. Even the mass chains at $A=257$, 258, 259 may not be produced in full measure. The onset of neutron-induced fission is delayed until $A\simeq 260$ only because the r process builds through neutron-rich nuclei for which the fission-inducing parameter Z^2/A is abnormally low, and also because the average neutron energy is moderately low ($kT\sim 100$ kev at $T\sim 10^9$ degrees). These last points constitute the central explanations underlying the difference between the fission decay curve of the californium fraction of the Bikini thermonuclear debris and the energy input curve expected for a Type I supernova. In the Bikini debris the 2.2-year Cf^{252} activity dominated the fission decay curve after about 200 days (Fi56) and indicated that 50 times as much Cf^{252} was produced as Cf^{254}. This factor of 50 is required because Cf^{252} decays only 3% of the time by spontaneous fission. In the thermonuclear test the capture path started with U^{238} and was limited to the neutron-rich isotopes of uranium ($Z=92$) even as it built through the progenitors of the beta-stable californium isotopes at 252 and 254. For the r process we see from Table VII,2 and Fig. V,2 that $Z=87$ and 88 at $A=252$ and $A=254$, respectively. Thus, the progenitors in the Bikini explosion had relatively higher Z and were more susceptible to neutron-induced fission as compared to neutron capture than are the progenitors in an r-process event. In addition, they were almost certainly subject to a neutron flux containing a high-energy component in the many Mev range and for this reason interacted relatively more often through fission than through capture. The result was certainly that the Bikini synthesis of the neutron-rich uranium isotopes (which ultimately decay to Cf^{252}, Cf^{254}, etc.) occurred only through an infrequent (1–10%) capture

event relative to fission at each mass number in the build-up from $A=238$ to 254. This competition with fission explains in a very plausible manner why, even though the high-energy neutron-capture cross section remained constant at the geometrical size of the neutron-rich isotopes of uranium, yet the relative synthesis must have decreased markedly with increasing A so as to have produced Cf^{254} only in the observed 2% of the amount of Cf^{252}. Thus, in the fission decay curve of the Bikini californium fraction, the Cf^{252} decay with its 2.2-year half-life should dominate, as it does, after about 200 days. This will not at all be the case in the r process where, as shown in Table VII,2, Cf^{252} and Cf^{254} are produced in the ratio of 0.097 to 0.139, essentially one to one.

We now return to the quantitative calculation of the radioactive energy decay curve which follows an r-process event. In Fig. VIII,2 is shown the energy release in ergs/day as a function of time for (i) the beta decays of nuclei with $A<70$, (ii) the alpha and beta decays with $A>70$, (iii) the spontaneous fission decay, and (iv) the total sum for these three cases. This figure is based on the calculations given in Table VIII,1, and these in turn are based on the abundance distribution for steady-state flow but without cycling from $A=110$ to 260. Figure VIII,2 thus represents the radioactive energy decay curve for an r-process event, such as might take place in a supernova outburst, which would have produced the atomic abundance distribution of Figs. VII,3 and VII,4 plus an extension to atomic weights below $A=70$. The main contributions to the energy release after the first few days through the first two years come from Ca^{47}, Fe^{59}, Ra^{228}, Cf^{252}, and Cf^{254}. These contributions are summarized in Table VIII,2.

Figure VIII,2 shows that the total energy release parallels the spontaneous fission release from 50 to 450 days. The spontaneous fission release during this period is almost entirely due to Cf^{254}. In the labels for each curve in Fig. VIII,2 we also include the total mass, in atomic mass units, involved in the production of each of the three types of decay given as (i), (ii), and (iii) above. Thus, the alpha- and beta-decay energy for $A>70$ in Mev/day (right-hand ordinate) is that for 2×10^4 atomic mass units of material which has been computed by multiplying the abundances of Table VII,2 by the appropriate atomic weights, A, and summing over $A>70$. In the left-hand ordinate we have obtained the energy release in ergs/day for the same numerical amount of material in grams by multiplying by 1.6×10^{-6} erg/Mev and dividing by 1.66×10^{-24} gram/atomic mass unit. The number of mass units producing the beta decay for $A<70$ is estimated to be 8×10^4 by extending the calculations of Table VII,3 in an approximate manner to $A<70$. Two estimates are given for the amount of mass associated with the spontaneous fission curve. If only the Cm^{250}, Cf^{254}, and Fm^{257} which actually undergo fission are counted, the total mass is some 120 atomic mass units. However, it is not possible to suppose that these fissionable products are produced alone and thus for another estimate we associate with them the production of all of the mass chains above $A=110$. This is the appropriate model for the situation discussed previously in which the neutron flux is great enough to move all of the material into the region $A=110$ to 260 and cycle it there. The yield from $A=110$ to 150 must be multiplied by one-half because only half of the fission fragments return to this region.

These considerations show that the energy release per gram of r product is much greater for fission than for alpha and beta decay. This is illustrated graphically in Fig. VIII,3 where the energy release is plotted for three different cases in which the same amount of material, namely, 1% of a solar mass (2×10^{31} g), is converted to r products. In Fig. VIII,3 the ordinate has been expressed in absolute magnitude with zero magnitude corresponding to 2.3×10^{40} ergs/day; this is done to facilitate comparison with actual supernova light curves which are shown and discussed in Sec. XII. A total

TABLE VIII,2. Principal radioactive decays of Table VIII,1.

A	Element	Longest half-life, days or years	Decay type	Decay energy Q Mev	Initial abund. n_0 (Si $=10^6$)	Total energy n_0Q Mev	Initial en. rate $0.7\ n_0Q/t_{1/2}$ Mev/day
33	P	25d	β	0.12	50	5.95	0.165
47	Ca	4.7d	β	1.83	100	183	26.9
59	Fe	45d	β	1.36	30	40.8	0.628
85	Kr	10.4y	β	0.31	3.4	1.05	1.9×10^{-4}
89	Sr	54d	β	0.66	3.4	2.24	0.0287
91	Y	58d	β	0.70	2.2	1.54	0.0183
95	Zr	65d	β	1.73	2.2	3.81	0.0408
131	I	8.05d	β	0.50	1.73	0.865	0.0745
140	Ba	12.8d	β	~2.4	0.45	~1.08	~0.058
144	Ce	285d	β	~1.5	0.45	~0.68	~0.0016
194	Os	~2y	β	1.01	0.88	0.889	$\sim8.4\times10^{-4}$
225	Ra	14.8d	$\alpha+\beta$	27.9	0.045	1.26	0.059
228	Ra	6.7y	$\alpha+\beta$	35.2	0.032	1.13	3.2×10^{-4}
250	Cm	7.5×10^3y	α, SF(75%)	170	0.139	23.6	5.97×10^{-6}
252	Cf	2.2y	α, SF(3%)	12.9	0.097	1.25	0.0011
254	Cf	61 or 56.2d	SF	220	0.139	30.6	0.349
257	Fm	~10d	α, SF(6%)	18.5	0.091	1.68	~0.116

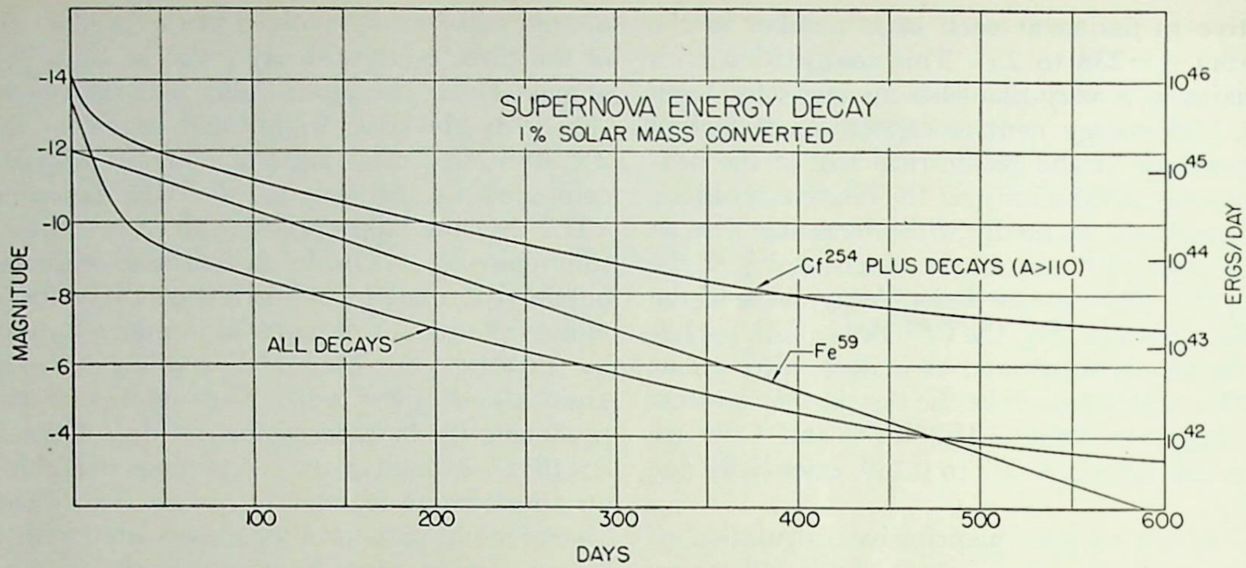


FIG. VIII,3. The radioactive energy release of the products of the r process calculated by assuming that one percent of a solar mass (2×10^{31} g) of these products is produced. Three cases are illustrated; (i) the energy release for all alpha, beta, and fission decays in relative abundances similar to those found for r-process products in the atomic abundance curve, (ii) the energy release for an extreme r-process event in which all of the converted material is moved into the region above $A=110$ and cycled there, and (iii) the energy release from Fe^{59}, when this isotope is produced in maximum quantity in a process in which only a few neutrons are made available per iron group nucleus. Compare with Fig. XII,1.

mass of one percent of a solar mass is taken simply as a reference point in this diagram. The total mass necessary to be converted to produce the observed light curves of supernovae is discussed for two specific cases, the Crab Nebula and the supernova in IC 4182, in Sec. XII.

The three cases which have been treated in Fig. VIII,3 are as follows. One case is the energy release for all activities in equilibrium with half-lives greater than 3 days, produced in an r process which would give an abundance distribution similar to that necessary to account for the r-process isotopes in the atomic abundance curve. This would occur, as indicated previously, if steady flow was achieved but no cycling occurred in the region above $A=110$. Another case illustrated is the energy release just for the activities produced above $A=110$. This is then the type of energy curve expected for a supernova in which the neutron flux was so great that all of the converted material was pushed into the region above $A=110$ and cycled there (one half cycling $A=110$ to 150). In this curve the Cf^{254} fission decay dominates the curve from 56 to 450 days but other activities take over after 450 days. The relation between this energy decay curve and the light curve of the supernova in IC 4182, which was observed for about 600 days after maximum, is also considered in Sec. XII.

The third case shown in Fig. VIII,3 is one calculated for the special case in which only a few neutrons are made available per iron group nucleus. In this case the most important activity produced is that of Fe^{59}, which has a half-life of 45 days. It is produced in maximum abundance when $n_c/Fe^{56}=3$, in which case 20% of the Fe^{56} is converted into Fe^{59} (Fo55a, Appendix II) and the remainder into nearby nuclei. The conditions under which the decay curve produced by Fe^{59} could explain an observed supernova light curve are discussed in Sec. XII, although from the astrophysical standpoint we do not consider this to be a very likely occurrence.

(2) *Abundance Yield in the r Process below A=70; the Production of Titanium; Discussion of Short-Period Radioactivity Following the r Process*

We have postponed to this point discussion of our estimate for the r-process yields below $A=70$. Two important activities, 4.8-day Ca^{47} and 45-day Fe^{59}, occur in this region. We have mentioned Fe^{59} above in discussion of a rather special possibility. Here we discuss its production in a steady-state r process similar to that discussed in Sec. VII when the abundances above $A=70$ were calculated. There is one important difference, however. When Fe^{56} is taken as the starting point (infinite source) in the r process, the first waiting point is at $A=68$, and no production below this atomic weight is achieved. Thus, the source material for building isotopes below $A\sim 70$ must be the light elements rather than the iron group elements. This leads to considerable uncertainty in estimating the r-process yields below $A=70$.

Another reason why detailed calculations have not been made on the r process below $A=70$ is the fact that the s, e, and α processes are the major contributors in this region. Only for S^{36}, Ca^{46}, and Ca^{48} in this region is there no other method of production than the r process. It is also possible that Ti^{47}, Ti^{49}, and Ti^{50} should be assigned to this process. All of these isotopes except Ca^{46} have abundances $\sim$100 (Su56). Thus empirical data in this region are lacking. Although good mass data are

available, we have hesitated to use our expression for the effective W_β for light nuclei in calculating τ_β and hence abundances, since it assumes a high density of levels in the daughter product. Since the abundance is proportional to W_β^{-5}, one can expect considerable fluctuations for the r-process products. This may explain the behavior illustrated by the abundance ratios $Ca^{46}:Ti^{47}:Ca^{48}=1.6:189:88$. An abundance peak for $N=28$ in the r process might be expected at somewhat lower A, and thus the abundance of Ca^{46} seems definitely low and must be attributed to a fluctuation in the value of W_β at the progenitor for $A=46$. Ti^{47} may also be produced in the s process or possibly in the e process. The abundance produced in the s process should be considerably less than the abundances of Sc^{45} and Ti^{46}, which are 63 and 194, respectively, since the s process is dropping rapidly in this region. It should also be produced less abundantly than $Ti^{46}(194)$, $Ti^{49}(134)$, and $Ti^{50}(130)$ in the e process. Thus it would seem reasonable to attribute to the r process a part of the mass-chain at $A=47$ comparable in abundance to Ca^{48} rather than to the abnormally low Ca^{46}. Ca^{47} is in this chain and we round off its abundance to about 100 as given in Table VIII,1. As indicated in Fig. VIII,2, this gives rather a large contribution in the early part of the energy decay curve. However, as discussed in Sec. XII, the early part of the supernova light curve may not be simply a radioactive energy decay curve, since it may represent cooling of the expanding envelope for a short period after the explosion.

The amount of production of the mass-chain at $A=59$, and of Fe^{59} in particular, is very difficult to estimate. We can expect that the nucleus produced directly in the r process might have $Z=20$. An atomic abundance estimate of 30 is a reasonable value. The energy release in the decay of the Fe^{59} is slightly larger than that in the decay of the Cf^{254} when the atomic abundance yield of the Fe^{59} is assumed to be 30. The abundances are in the ratio $30/0.139\simeq 220$ while the energy release ratio is $1.35/220=1/160$. The Ca^{47} energy release is ~ 5 times that of Cf^{254}. This, of course, will not be the case in a supernova in which a large ratio of neutrons to iron is available and only activities with $A>110$ are produced. However, this illustrates the fact that in any case the energy release in the Cf^{254} is only a small part of the *total* energy release from the radioactivity following a supernova explosion. In each mass-chain there will be very rapid beta decays releasing some 10 Mev in beta-ray energy. We estimate that the Cf^{254} is only about 0.05 to 0.5% by number of all the nuclei produced in the r process. Thus the total energy released radioactively is 10 to 100 times that from Cf^{254} but is emitted within a matter of a few minutes or hours of the initial explosion. The main point is this: If the Cf^{254} release is $\sim 10^{47}$ ergs, which is a reasonable estimate, then the total radioactive energy release will be 10^{48} to 10^{49} ergs. This is indeed a large value, but it does not exceed the total energy available, some 10^{50} ergs, from the hydrogen and helium reactions which take place in the envelope of the supernova.

(3) *Abundance of Fission Fragments Produced in the r Process*

At this point a remark concerning the contribution of fission fragments to the atomic abundance distribution is in order. Termination of the r process by very rapid spontaneous fission and neutron-induced fission during the supernova explosion will occur at $A\sim 260$, $Z\sim 90$. Fission of such neutron-rich nuclei will probably result in the production of more neutrons than are observed for beta-stable nuclei; we estimate ~ 6. On the assumption that the fission is asymmetrical in charge and mass we then expect fragments which are spread about $A\sim 108$, $Z\sim 38$, and $A\sim 146$, $Z\sim 52$, as shown diagrammatically in Fig. V,2, although we have rounded off these numbers to $A=110$ and 150 in preceding discussions. The neutron-rich fragments are very similar to the nuclei involved in the capture path and are swept up into the r process and do not produce any particular effects near $A\sim 108$ or 146. They tend to cycle material above $A\sim 108$ and to establish steady-state abundances in this region, even though they are far from the iron group elements and light elements with which the synthesis starts. The neutrons released in the fission are recaptured as the r process proceeds.

On the other hand, the ultimate decay by spontaneous fission of Cm^{250}, Cf^{252}, Cf^{254}, and Fm^{257} produces a total yield of both light and heavy fragments equal to $0.75\times 0.139+0.03\times 0.097+0.139+0.06\times 0.091=0.252$ on the $Si=10^6$ scale (see Tables VII,2 and VIII,1). This abundance is spread out over the usual distribution in fragments and the peak yield can be expected to be about 8% of the total or about 0.020. This is illustrated graphically in Fig. VII,3, where it will be noted that the fission fragment abundance is small compared to that produced directly in the r process.

C. Age of the Elements and of the Galaxy

Transuranium elements are produced by the r process. Those with atomic weights of the form $235+4m$, $m=0, 1, 2, 3\cdots$ are progenitors of U^{235} provided alpha decay is more probable than fission. From Table VIII,1 this appears to be the case for $m\leq 5$ but for $m>5$ it seems that fission must dominate. Thus there are six progenitors of U^{235} at atomic weights 235, 239, 243, 247, 251, and 255. The number of progenitors of U^{238}, of atomic weights $238+4m$, is even more restricted. As can be seen in Table VIII,1 the beta-stable nuclei at $A=238$, 242, and 246 decay by alpha emission while that at $A=250$ decays 25% of the time by alpha emission and 75% by fission. Starting with Cf^{254}, nuclei of atomic weights 254, 258, etc., decay predominantly by fission. This information is summarized in Table VIII,3, where we also list the progenitors of the other parents of the radioactive series, Th^{232} and Np^{237}.

TABLE VIII,3. Production of parents of radioactive series; progenitors decaying by alpha emission rather than spontaneous fission.

Half-life	$_{90}Th^{232}(4n)$ $1.39\times10^{10}y$		$_{93}Np^{237}(4n+1)$ $2.2\times10^{6}y$		$_{92}U^{238}(4n+2)$ $4.51\times10^{9}y$		$_{92}U^{235}(4n+3)$ $7.13\times10^{8}y$	
	A	Yield	A	Yield	A	Yield	A	Yield
	232	0.084	237	0.057	238	0.110	235	0.057
	236	0.057	241	0.221	242	0.157	239	0.110
	240	0.110	245	0.103	246	0.103	243	0.157
	244	0.157	249	0.139	250(25%)	0.035	247	0.103
	248	0.139	253	0.097			251	0.097
	252(97%)	0.094					255	0.139
Total yield		0.641		0.617		0.405		0.663

Ratio $U^{235}/U^{238}=0.663/0.405=1.64$
$Th^{232}/U^{238}=0.641/0.405=1.58$

If the abundances produced by the *r* process were the same for all progenitors, the ultimate abundance of U^{235}, after alpha and beta decays at 239, 243, etc., would exceed that of U^{238} by a factor of $6/3.25=1.85$. Calculations discussed in the previous sections make it possible to improve upon this factor by totaling the yield of the progenitors for the individual radioactive parents, U^{235}, U^{238}, and Th^{232}, Np^{237} as well. Yields for the progenitors and the totals after alpha and beta decay are given in Table VIII,3. The ratio of U^{235}/U^{238} produced by the *r* process and subsequent decay is 1.64, slightly less than the ratio of the number of progenitors. This is to be expected since the even-A progenitors of U^{238} will individually be produced in greater abundance than the odd-A progenitors of U^{235} because of the even-odd pairing term which decreases the beta-decay energy of even-A waiting points and thus increases even-A waiting point abundances. This effect is most clearly demonstrated by the calculations in Sec. VII which show peaks at $A=130$ and 194, where each A is a waiting point. We digress, in order to discuss this effect more fully.

For the $A=130$ peak, results of the calculations given in Table VII,2 are that $W_\beta=11.46$, 9.30, and 9.61 Mev at $A=125$, 126, and 127, respectively. For $A=126$ the value would have been $W_\beta=10.50$, i.e., it would have been intermediate between the values for 125 and 127, if we had not subtracted $\delta_A/2=1.20$ Mev from this value. The calculated abundances were then found to be 0.18, 0.50, and 0.42, and these values must be compared with the observed values for Te^{125}, Te^{126}, and I^{127}, i.e., 0.328, 0.874, and 0.80. The computed values are low but are relatively correct. Empirically, we can estimate the even-odd pairing effect from $2Te^{126}/(Te^{125}+I^{127})=1.55$. The main uncertainty in this calculation arises from the uncertainty in the abundance ratio of tellurium to iodine.

Three osmium isotopes are produced in the rising portion of the $A=194$ peak. Using $\delta_A/2=0.78$ Mev, we find that $W_\beta=13.06$, 13.14, and 11.54 Mev at $A=188$, 189, and 190, respectively. The predicted abundances are then 0.09, 0.09, and 0.17, and these are in good agreement with the observed abundances of Os^{188}, Os^{189}, and Os^{190}, which are 0.133, 0.161, and 0.264, respectively. In this case, on the basis of relative isotopic abundances alone, the even-odd pairing term is given by $(Os^{190}+Os^{188})/(2\times Os^{189})=1.23$. Incidentally, the observed abundance ratio of the two osmium isotopes $Os^{190}/Os^{188}=1.98$ represents a substantial check on our calculated abundances (proportional to W_β^{-5}), which give a ratio of 1.85. This agreement indicates that our choice of parameters B_A and Z_A and our use of $W_\beta=B_A(Z_A-Z-2.5)+$pairing term has some empirical validity. Off the abundance peaks, the even-odd effects are not so clear cut in our method of calculation, since odd-A nuclei can result from even-A waiting points and vice versa. However, there is some correlation along even and odd lines as is clear from a perusal of Table VII,2.

An even-odd effect of 1.13 yields the change from 1.85 to 1.64 discussed above for the uranium isotopes, and this seems to be a reasonable value. It neglects the effect of delayed neutron emission from excited states of the nuclei in the decay of the various mass chains to their stable isobars. This is a small effect, but does give some transfer of even-A material to odd-A material since the most highly excited states result in beta transitions from odd Z-odd N-even A nuclei to even Z-even N-even A nuclei. If the even-odd effect were as high as 1.55, as it is at the $A=130$ peak, the production ratio would be 1.20. By a selective choice of parameters entering into our calculation the ratio could be reduced in extremum to a value as low as unity. It is our considered opinion that it is highly unlikely for the ratio to be less than this value. A reasonable value for an upper limit would be 2.

With these estimates of the production ratio of U^{235}/U^{238} in the *r* process, we are now able to calculate the age of these uranium isotopes (and by inference of the other *r*-process isotopes) using their presently observed terrestrial abundance ratio.

(i) Let us suppose that all of the uranium were produced in a single *r*-process event (e.g., one supernova) at one moment of time, $t_0\times10^9$ years ago. Then, using a present-day ratio of $U^{235}/U^{238}=0.0072$, and using half-lives of 7.13×10^8 years and 4.51×10^9 years for U^{235} and U^{238}, respectively, we have the following equa-

tion for the abundance ratio U^{235}/U^{238} as a function of t_0

$$\frac{U^{235}(t_0)}{U^{238}(t_0)}=0.0072\times 2^{t_0[(1/0.713)-(1/4.51)]}=0.0072\times 2^{1.18t_0}.$$

This relationship is plotted in Fig. VIII,4. At the formation of the solar system 4.55×10^9 years ago $U^{235}/U^{238}=0.29$, which is definitely below the lower limit of unity for the production ratio discussed above. We find $U^{235}/U^{238}=1$, 1.64, and 2 at $t_0=6.0$, 6.6, and 6.9×10^9 years, respectively. Our best value for the time of a single-event synthesis of U^{235} and U^{238} is thus 6.6×10^9 years ago. If these isotopes were produced in a single supernova, this is its date. For each change of a factor of two in the production ratio of the uranium isotopes, this figure changes by 0.85×10^9 years.

There is some confirmatory evidence for the above calculations in the Th^{232}/U^{238} abundance ratio, which is, however, much more uncertain than the U^{235}/U^{238} ratio. Suess and Urey estimate that the present-day ratio $Th^{232}/U^{238}=3$ to 3.5. Our production ratio is $0.641/0.405=1.58$. From these data we find $t_0=6.2$ to 7.7×10^9 years. This time estimate is obviously quite sensitive to the value taken for the present thorium to uranium ratio but is not inconsistent with the value of t_0 obtained from the U^{235}/U^{238} ratio.

(ii) Alternatively, let us consider the problem on the basis that production of uranium started a time $t_0\times10^9$ years ago and that, instead of production being confined to just one moment, production took place at a constant rate until a time $t_1\times10^9$ years ago, after which no more production occurred. This model corresponds to production in a series of supernovae with uniform mixing of their products over the interval t_0 to t_1. Let $U^{235}(t)$ and $U^{238}(t)$ be the abundances at any time, t, and let λ_{235} and λ_{238} be the production rates taken as independent of time. If t is taken as the time measured backward from the present epoch, we then have

$$\left.\begin{aligned}-\frac{dU^{235}(t)}{dt}&=\lambda_{235}-\frac{0.693}{0.713}U^{235}(t)\\ -\frac{dU^{238}(t)}{dt}&=\lambda_{238}-\frac{0.693}{4.51}U^{238}(t)\end{aligned}\right\}\quad t_0\geq t\geq t_1.$$

These equations taken together with the boundary conditions

$$U^{235}(t_0)=U^{238}(t_0)=0$$

integrate to give

$$\frac{U^{235}(t)}{U^{238}(t)}=\frac{0.713\lambda_{235}[1-2^{-[(t_0-t)/0.713]}]}{4.51\lambda_{238}[1-2^{-[(t_0-t)/4.51]}]}\quad t_0\geq t\geq t_1.$$

This result, together with the condition that the free decay of U^{235} and U^{238} from t_1 to the present day must yield a ratio equal to 0.0072, determines the following relation between t_1 and t_0

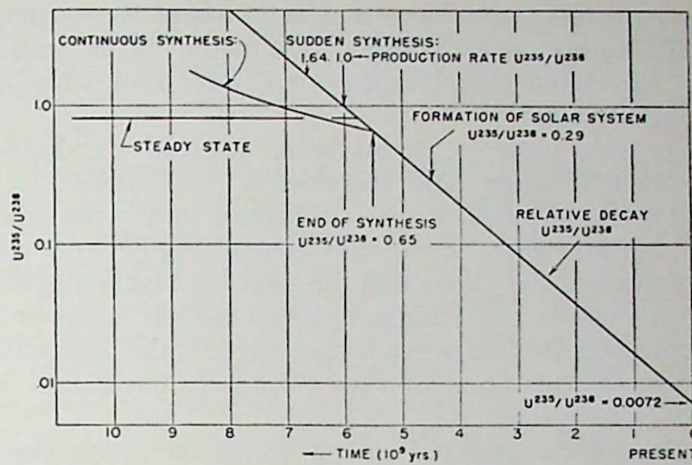


FIG. VIII,4. The ratio U^{235}/U^{238} as a function of time. This curve has been employed to determine the age of terrestrial uranium by the method described in Sec. VIII C.

$$\frac{U^{235}(0)}{U^{238}(0)}=\frac{0.713\lambda_{235}[1-2^{-[(t_0-t)/0.713]}]2^{-t_1/0.713}}{4.51\lambda_{238}[1-2^{-[(t_0-t)/4.51]}]2^{-t_1/4.51}}=0.0072.$$

With

$$\lambda_{235}/\lambda_{238}=1.64,$$

this reduces to

$$\frac{[1-2^{-[(t_0-t_1)/0.713]}]}{[1-2^{-[(t_0-t_1)/4.51]}]}\times2^{-1.18t_1}=\frac{0.0072\times4.51}{1.64\times0.713}=0.0278.$$

The values of t_0 corresponding to a set of values of t_1 are given in the following tabulation, along with values of $U^{235}(t_1)/U^{238}(t_1)$:

t_1 (in 10^9 years)	4.5	5.0	5.5	6.0	6.6
t_0 (in 10^9 years)	18	11.5	8.5	7.4	6.6
$U^{235}(t_1)/U^{238}(t_1)$	0.29	0.43	0.65	0.98	1.64

The computed value for t_0 is very sensitive to t_1 if t_1 is chosen near the time of formation of the solar system at about 4.5×10^9 years ago. On the other hand, t_0 does not change rapidly once t_1 exceeds 5.5×10^9 years ago, since t_0 must equal t_1 at 6.6×10^9 years, the computed time for a single-event synthesis. Evidence on the time interval between nucleogenesis and the formation of the meteorites of the solar system has been given by Wasserburg and Hayden (Wa55b). Following Suess and Inghram they argue that Xe^{129} would be abnormally abundant in meteorites if the meteorites crystallized before I^{129}, with an observed half-life of 1.7×10^7 years, had all decayed. They did not find abnormal amounts of Xe^{129} in the Beardsley chondrite and concluded that it crystallized at least 23.7 half-lives or 4.1×10^8 years after the production of the I^{129}. In their argument they assume that I^{129} and I^{127} were produced in equal amounts. This is consistent with our point of view that I^{129} is a progenitor of Xe^{129} and Suess and Urey's abundance ratio $Xe^{129}/I^{127}=1.05/0.80=1.30$. Our calculations in Table VII,2 yield $I^{129}/I^{127}=1.17/0.42\cong2.8$ in

the *r* process. Actually, the calculations of Wasserburg and Hayden are quite insensitive to the I^{129}/I^{127} ratio, our calculated value raising their lower limit only to 4.3×10^8 years.

On this basis it would seem justified to choose $t_1 \geq 5.0\times10^9$ years, and therefore t_0 lies between 11.5 and 6.6×10^9 years. For nucleogenesis in a continuous and uniform series of supernovae occurring between the formation of the galaxy and the time of formation of the cloud of gas or protostar from which the sun condensed, we can take 10^{10} years as a round figure for the beginning epoch. On this point of view the age of our galaxy is thus $\sim10^{10}$ years.

There is one point of view in which the above ideas must be somewhat modified. In the steady-state cosmologies of Hoyle (Ho49) and of Bondi and Gold (Bo48) it may be expected that the U^{235}/U^{238} ratio in a sufficiently large volume will remain constant due to supernova activity, in spite of the apparent universal expansion and the constant dilution by the addition of matter in the form of hydrogen which is necessitated in order to keep the mass energy content of the specified volume at a fixed value. The characteristic time involved for the removal by expansion of one-half of a particular type of matter from a fixed volume is $0.693T/3$, where T is the reciprocal of Hubble's constant. Taking T as 5.4×10^9 years we have for our characteristic half-life 1.25×10^9 years. The decay lifetime for any radioactive material must be properly corrected for this value. The effective lifetimes of U^{235} and U^{238} become 0.45×10^9 and 0.98×10^9 years, respectively.

On the steady-state point of view $t_0=\infty$, and after a long period of production one would expect that at time t_1, after which no further contribution to solar system material occurred, we should have

$$\frac{U^{235}(t_1)}{U^{238}(t_1)}=\frac{0.45\lambda_{235}}{0.98\lambda_{238}}=0.75.$$

This is just the value expected from the present-day ratio of $U^{235}/U^{238}=0.0072$ now if the solar system material has been isolated for some 5.7×10^9 years. We may conclude that the present abundance ratio of the uranium isotopes is consistent with the steady-state cosmology if the method of production of their isotopes is the *r* process as described. We do not wish to overemphasize this conclusion, but rather to emphasize that studies concerning the origin of the radioactive elements may lead to objective tests of the various cosmological theories.

It can also be noted that there exists an approximate indication of the epoch of synthesis by the *s* process. The radioactive isotope K^{40} is in the main line of the *s* process. Figure VI,1 shows that for this isotope $\sigma N\sim10^4$ for σ in mb and N on the scale $Si=10^6$. We estimate in the appendix that $\sigma\sim100$ mb so at the time of K^{40} production $N\sim100$. It is now, $N=0.38$. Using the known K^{40} half-life of 1.3×10^9 years we find that the K^{40} was produced $\sim10^{10}$ years ago. This calculation could be made with great assurance if the capture cross sections for 15-kev neutrons were measured accurately for K^{39}, K^{40}, and K^{41}. Our calculation assumes the K^{40} cross section to be 12.5 times that of K^{39} which has a closed shell of neutrons.[20] Even if this factor has the very high value of 100 the epoch for K^{40} formation will be reduced only to 6.5×10^9 years ago.

In any case we consider the minimum age of the isotopes of uranium, namely 6.6×10^9 years with an uncertainty on the low side of at most 0.6×10^9 years, to be significantly greater than the age of the solar system, 4.5×10^9 years. This age is indeed comparable to the ages, 6.5×10^9 years, of the oldest known clusters in the galaxy (Jo56, Ho55, Ha56c). Since our age is a minimum for that of the elements and thus, on the hypothesis of stellar nucleogenesis, for the galaxy, it would not be unexpected if still older objects were eventually found in the galaxy.

D. Termination of the *s* Process; the Abundances of Lead, Bismuth, Thorium, and Uranium

The *s* process is terminated at Bi^{209} by the onset of natural alpha-particle radioactivity. Detailed discussion of the termination has been postponed to this point so that it can be applied, along with the discussion of the *r*-process termination, to calculation of the abundances of lead, bismuth, thorium, and uranium, which can be expected to be produced in the *s* and *r* processes.

As illustrated in Fig. VIII,5, the *s* process proceeds through the mercury isotopes to stable Tl^{203} and thence by neutron capture to Tl^{204} which decays 98% by beta emission to Pb^{204} and 2% by electron capture to Hg^{204}. We can neglect this last branching. The half-life of Tl^{204} is 4.1 years, which we shall take to be short compared to its mean neutron-capture time. At Pb^{204} neutron capture leads to Pb^{205} which decays by electron capture with a lifetime of approximately 5×10^7 years to Tl^{205}.

Pb^{205} is essentially stable on the time-scale of individual captures in the *s* process, and the neutron capture continues through the stable lead isotopes Pb^{206}, Pb^{207}, and Pb^{208} to beta-active Pb^{209}. After the *s* process finally ends, the Pb^{205} produced in the capture chain eventually does decay to Tl^{205}. Pb^{209} decays in 3.3 hours to stable Bi^{209} which, upon neutron capture, forms the 2.6×10^6-year alpha-particle-emitting ground state of Bi^{210} or the 5.0-day electron-emitting isomeric state, RaE. For slow neutrons the capture results 44% of the time in the ground state and 56% of the time in the isomeric state. It is reasonable to assume that this branching ratio will hold for the 15-kev neutrons involved in the *s* process. The RaE decays quickly before neutron capture to Po^{210} which is an alpha-particle emitter with 138.4-day half-life, fast compared to neutron capture times. Thus Pb^{206} is

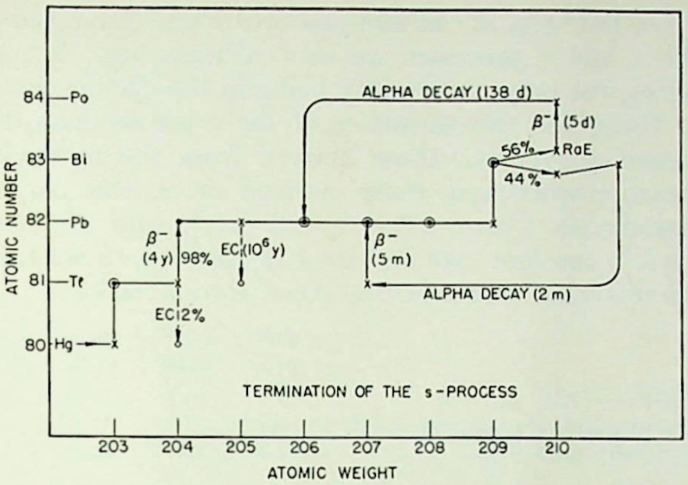


FIG. VIII,5. This diagram shows the details of the termination of the *s* process. The half-life for electron capture in Pb^{205} should be changed to $\sim 5\times 10^7$ years.

formed and this branch of the termination of the *s* process can result in a cycling process between Pb^{206} and Bi^{210}. The long half-life of the ground state of Bi^{210} means that it captures a neutron before decaying, so that in this case the termination is delayed until Bi^{211} is formed; this then decays rapidly through Tl^{207} to Pb^{207}, which can then cycle back through lead and bismuth. The result is a steady production of Pb^{206}, Pb^{207}, Pb^{208}, and Bi^{209} by the *s* process. The steady-flow relative abundances of these nuclei are then given by

$$an_{204}\sigma_{204}=an_{205}\sigma_{205}=1.8n_{206}\sigma_{206}$$
$$=n_{207}\sigma_{207}=n_{208}\sigma_{208}=n_{209}\sigma_{209},$$

where the σ's represent neutron capture cross sections, a is the factor for the number of terminal cycles, and the factor $1.8=1+44/56$ arises because Pb^{206} is replenished in only 56% of the cycling processes.

The capture cross section for 15-kev neutrons have been discussed in Sec. V and are given in the appendix. These cross sections are 100 and 50 mb for the odd isotopes Pb^{205} and Pb^{207}, 25 mb for the even isotope Pb^{206}, 10 mb for the even doubly-magic isotope Pb^{208}, and 15 mb for the odd magic Bi^{209}. These values are consistent with the energies gained on neutron addition, i.e., 8.10, 7.38, 6.73, 3.87, and 4.67 Mev, respectively. There is a problem concerning the neutron addition energy and the capture cross section for Pb^{204}. The reactivity measurements do not distinguish between even isotopes. From Huizenga's masses (Hu55) the energy is 6.80 Mev but has a considerably larger error than that for the other lead isotopes. It can be argued that Pb^{204}, with fewer neutrons than Pb^{206}, should more readily capture neutrons than Pb^{206}, and so the Pb^{204} cross section has been estimated as 50 mb. On this basis one-half as much Pb^{204} should be produced on the *s* process as Pb^{206}, as long as cycling is neglected. To see whether such a ratio is reasonable, we may compare it with the abundance ratios of other pairs of even isotopes which have been made in a similar way, i.e., by steady flow in the *s* process without cycling. We find, for example, that the abundance ratios $Te^{122}/Te^{124}=0.52$, $Xe^{128}/Xe^{130}=0.47$, $Ba^{134}/Ba^{136}=0.31$, and $Sm^{148}/Sm^{150}=1.52$.

An alternative way of estimating the relative lead capture cross sections is by comparison with the barium capture cross sections deduced from the relative abundances of the barium isotopes 134, 135, 136, 137, and 138. These lead to the completion of the $N=82$ neutron shell just as Pb^{204} to Pb^{208} lead to the completion of the $N=126$ shell. Ba^{134} and Ba^{136} are shielded, that is, they are made only in the *s* process, Ba^{138} is magic in the *s* process and all of its yield can be attributed to this process, while for Ba^{135} and Ba^{137} we have attributed two thirds of their abundances to the *s* process.

Thus, on the basis of our two independent estimates, the relative abundances of the lead isotopes produced in the *s* process are as follows:

	Pb^{204}	Pb^{205} (Tl^{205})	Pb^{206}	Pb^{207}	Pb^{208}	Bi^{209}
From reactivity σ	1	0.5	$1.1a$[a]	a	$5a$	$3.3a$
From barium abundances	1	1.8	$1.8a$[b]	$3.1a$	$29.6a$	...

[a] In case $a=1$, use 2 in place of $1.1a$.
[b] In case $a=1$, use 3.2 in place of $1.8a$.

Having discussed the relative abundances of the lead isotopes and bismuth expected to be produced by the *s* process, we must now calculate the relative abundance production expected in the *r* process. This can be done for Pb^{206}, for example, by first adding the abundances given in Table VII,2 for the mass chains 206, 210, 214, . . . 234, which decay fairly rapidly to Pb^{206}. For Pb^{206} this sum is 0.48 on our customary scale. If we wish to compare our results with the "nonradiogenic" lead isotope abundances at the time of formation of the solar system, we must add to this sum the amount of Pb^{206} resulting from the decay of U^{238} in the period between element production and solar system formation. This value depends somewhat on whether the element formation was continuous, or occurred in a single event, but for simplicity we have chosen the latter alternative. The decay period is then $(6.6-4.5)\times 10^9=2.1\times 10^9$ years, so that 27% of the U^{238} will have decayed. From Table VIII,3 we see that this adds 0.11 to the Pb^{206} abundance, bringing the total to 0.59. These abundances are summarized in Table VIII,4 which also includes the results of similar calculations for Pb^{207}, Pb^{208}, and Bi^{209}. Since Np^{237} decays rapidly as compared with the time scales involved here, the Bi^{209} abundance is just the sum over all mass chains 209, 213 . . . 253.

For the nonradiogenic lead isotope abundances it is customary to take the isotopic abundances observed in iron meteorites, which have the lowest observed proportions of radiogenic leads, and which are assumed to have crystallized without including parents of the radioactive series [see discussion on p. 68 of (Su56)]. For the present-day average for lead, Suess and Urey

recommend the chondritic meteorite ratios. According to Patterson *et al.* (Pa53, Pa55, Pa55a) the ratios are

	Pb^{206}/Pb^{204}	Pb^{207}/Pb^{204}	Pb^{208}/Pb^{204}	Total Pb/Pb^{204}
Iron meteorites, 4.5×10^9 yr ago	9.5	10.3	29.2	50.0
Chondrites "now"	19.4	15.9	38.6	74.9

Now Pb^{204} is not produced directly in the r process, since it is shielded by Hg^{204} and it does not result from any radioactive decay of r products. From the capture cross sections in the s process it may be expected that it will be produced to a lesser extent than the other isotopes, and this will be especially true if the s process does not just terminate at Bi^{209}, but cycles to some extent, thus building up Pb^{206}, Pb^{207}, and Pb^{208}. Thus the observed ratios given above are entirely reasonable on the basis of s process and r process synthesis of the elements.

On the other hand, these processes will have produced much more lead than commonly accepted in abundance estimates. This can be seen as follows. Pb^{204} is produced in the regular flow of the s process, and should have the corresponding abundance, which may be calculated as follows. The average value of σN, read off the curve in Fig. VI,3, is about 10 in the vicinity of $A=204$. The estimate given above for the cross section is 50 mb, and hence $N=0.20$. The cross-section estimate is probably high rather than low, so this abundance is, if anything, a lower limit. This leads to a total lead abundance of 10 for the iron meteorites and 15 for the chondrites. These values are much larger, for example, than those given by Suess and Urey, namely 0.31 and 0.47, respectively. On the other hand, our values are in good agreement with the solar abundance obtained by Goldberg *et al.* (Go57), i.e., a value of about 20 from the lead/hydrogen ratio and 10 from the lead/silicon ratio on the silicon $=10^6$ scale. Our view is that the meteoritic abundance of lead may be exceptionally low due to a fractionating process that took place when the meteorites were formed.

TABLE VIII,4. Production of Pb^{206}, Pb^{207}, Pb^{208}, and Bi^{209} in the r process.

	Pb^{206}	Pb^{207}	Pb^{208}	Bi^{209}
Long-lived parent	U^{238}	U^{235}	Th^{232}	(Np^{237})
Parent:				
Originally	0.405	0.663	0.641	0.617
4.5×10^9 years ago	0.295	0.083	0.577	0.0
Now	0.146	0.001	0.462	0.0
From long-lived parent:				
4.5×10^9 years ago	0.110	0.580	0.064	0.617
Now (no fractionation)	0.259	0.662	0.179	0.617
Short-lived progenitors	0.477	0.369	0.266	0.317
Total:				
4.5×10^9 years ago	0.587	0.949	0.330	0.934
Now (no fractionation)	0.736	1.031	0.445	0.934
Radiogenic	0.149	0.082	0.115	0.0

On the basis of our analysis of the lead produced in the s and r processes we can calculate lead isotope ratios, the only uncertainty being in the cycling factor a. Using the two estimates of the cross sections discussed above, i.e., those derived from the reactivity measurements and those derived from the barium abundances [with $\sigma(Pb^{204})=50$ mb], and choosing $a=6$ in the first case and $a=1$ in the second, we have the following values for the three isotopic ratios:

	Pb^{206}/Pb^{204}	Pb^{207}/Pb^{204}	Pb^{208}/Pb^{204}	Bi^{209}/Pb^{204}
From reactivity $\sigma(a=6)$	9.5	10.7	31.6	24.7
From barium abundances $(a=1)$	6.2	7.8	31.2	...
Iron meteorites (obs)	9.5	10.3	29.2	...

In this calculation we have taken into account the amount of lead produced by the decay of r-process isotopes before the freezing of the material of the meteorites 4.5×10^9 years ago. Thus for the first case $(a=6)$, $Pb^{206}/Pb^{204}=1.1\times6+0.587/0.2=9.5$. There is little to choose between the two sets of calculations, but we are inclined to favor the first since it would be rather accidental if the s process terminated at bismuth without some cycling at its termination.

In addition to obtaining a high abundance of lead through these calculations, we also find that the abundance of bismuth is high. Using the reactivity cross sections and the cycling factor $a=6$, we find that Bi^{209} should be produced with an abundance 20 times that of Pb^{204} in the s process. An additional contribution of $0.934/0.2=4.7$ comes from the r process (Table VIII,4). Thus, Bi^{209} should have an abundance of $24.7\times0.2=4.94$ on the silicon $=10^6$ scale whereas Suess and Urey, using a value due to the Noddacks, give 0.144.

On the other hand, our calculation of the abundances of the isotopes of thallium is in fair agreement with the abundances given by Suess and Urey. Thus Tl^{203}, with a cross section of 276 mb, has a calculated s-process abundance of $10/276=0.036$, to which must be added 0.057 from the r process, yielding a total of 0.093 which is only a factor of 3 higher than the Suess and Urey value of 0.0319. Tl^{205} results from the decay of Pb^{205} produced in the s process and thus the Pb^{205} cross section of 100 mb must be used in calculating its s-process abundance, i.e., $10/100=0.10$. Adding 0.057 for the r-process abundance, we obtain 0.157 which is about twice the value of 0.0761 given by Suess and Urey.

Our values for the mercury isotopes, as compared with the values of Suess and Urey, are shown in the following:

	s process	r process	Total	Suess and Urey
Hg^{196}	0.0	0.0	(p process)→	0.00045
Hg^{198}	0.042	0.057	0.099	0.0285
Hg^{199}	0.015	0.057	0.072	0.0481
Hg^{200}	0.042	0.057	0.099	0.0656
Hg^{201}	0.015	0.057	0.072	0.0375
Hg^{202}	0.042	0.057	0.099	0.0844
Hg^{204}	0.000	0.057	0.057	0.0194
			0.498	0.284

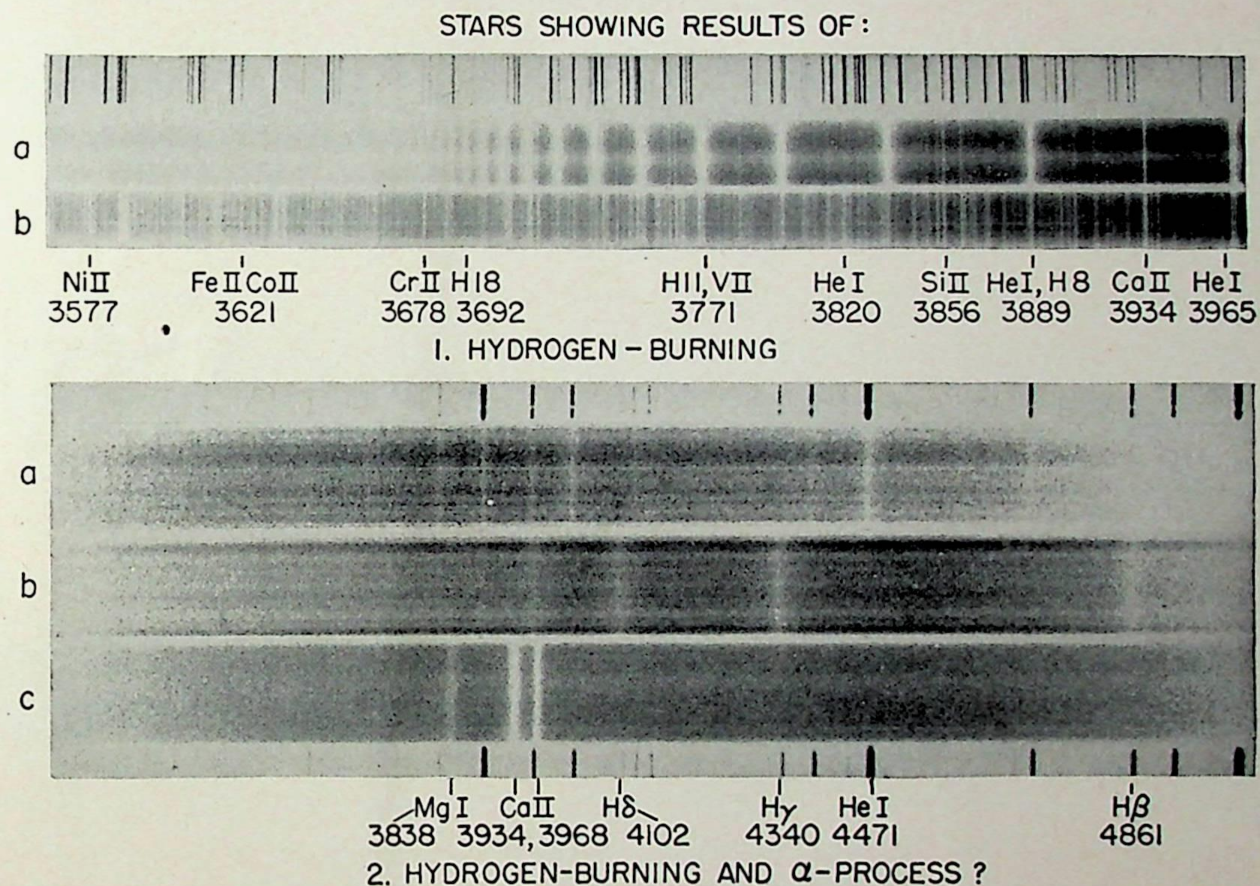


PLATE I.

PLATE 1. Portions of the spectra of stars showing the results of hydrogen burning and possibly the α process. Upper: (a) Normal A-type star, η Leonis, showing strong Balmer lines of hydrogen and a strong Balmer discontinuity at the series limit. (b) Peculiar star, υ Sagittarii, in which hydrogen has a much smaller abundance than normal. Lower: (a) White dwarf, L 1573–31, in which hydrogen is apparently absent. The comparison spectrum above the star is of a helium discharge tube; note the lines of helium in the star's spectrum. (b) White dwarf, L 770–3, which shows broad lines due to hydrogen only, for comparison with (a) and (c). (c) White dwarf, Ross 640, which shows only the two lines due to Ca II and a feature due to Mg I. All the spectrograms in this plate were obtained by J. L. Greenstein; the upper two are McDonald Observatory plates and the lower three are Palomar Observatory plates.

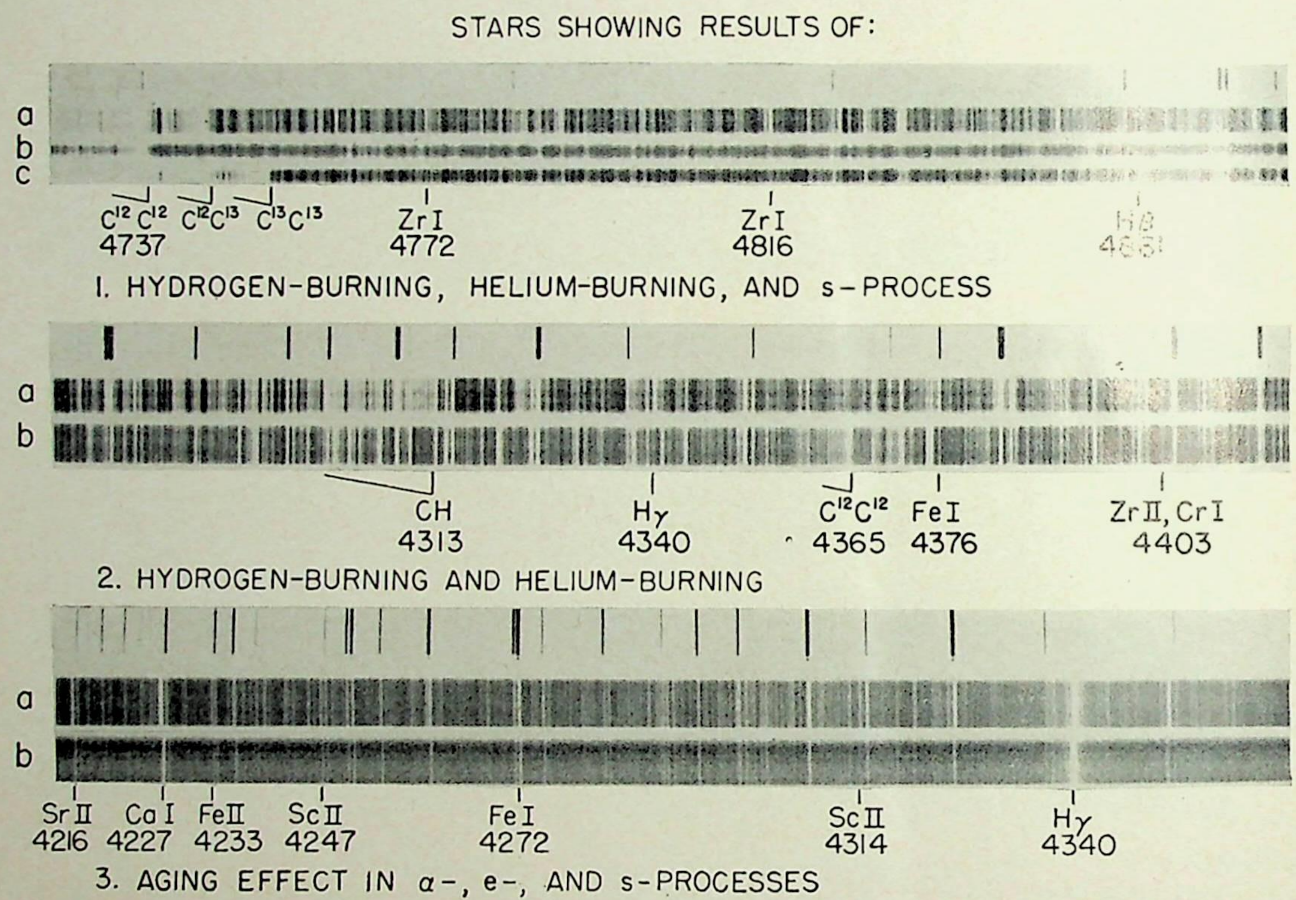


PLATE 2.

PLATE 2. Portions of the spectra of stars showing different aspects of element synthesis. Upper: (a) Normal carbon star, X Cancri, which has $C^{12}/C^{13} \sim 3$ or 4. (b) Peculiar carbon star, HD 137613, which shows no C^{13} bands, and in which hydrogen is apparently weak. (c) Normal carbon star, HD 52432, which has $C^{12}/C^{13} \sim 3$ or 4. Note that ZrI lines appear to be strongest in (a). Middle: (a) Normal carbon star, HD 156074, showing the CH band and Hγ. (b) Peculiar carbon star, HD 182040, in which CH is not seen, although the weak band of C_2 at λ 4365 is visible. Hγ is also very weak, indicating that hydrogen has a low abundance. Lower: (a) Normal *F*-type star, ξ Pegasi. (b) Peculiar star, HD 19445, which has a slightly lower temperature than ξ Pegasi, yet all lines but hydrogen are much weakened, showing that the abundances of α-, *e*-, and *s*-process elements are much lower than normal ("aging" effect). The middle two spectra were obtained by J. L. Greenstein, the remainder, with the exception of HD 137613, by E. M. and G. R. Burbidge.

STARS SHOWING RESULTS OF s-PROCESS

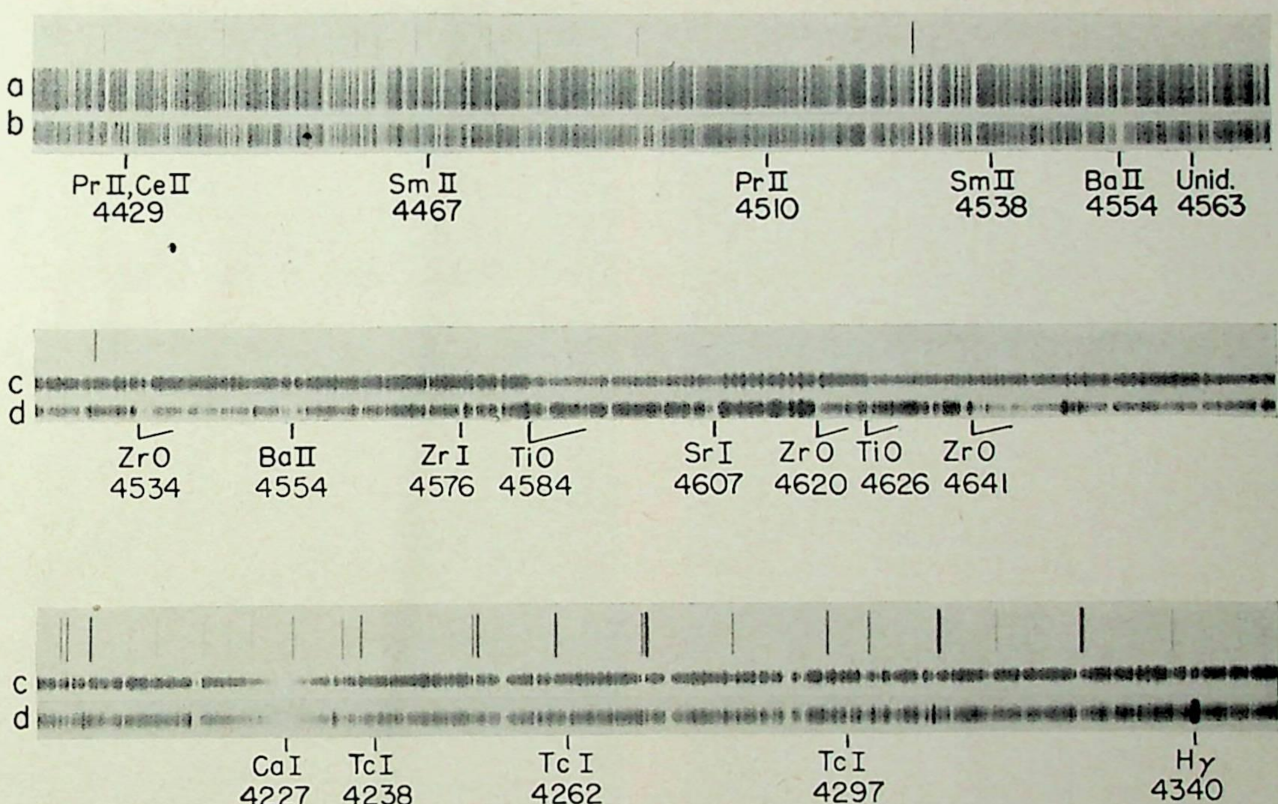


PLATE 3.

PLATE 3. Portions of the spectra of stars showing the results of the *s* process. Upper: (a) Normal *G*-type star, *κ* Geminorum. (b) Ba II star, HD 46407, showing the strengthening of the lines due to the *s*-process elements barium and some rare earths. Middle: (c) *M*-type star, 56 Leonis, showing TiO bands at λλ 4584 and 4626. (d) *S*-type star, R Andromedae, showing ZrO bands which replace the TiO bands. Lines due to Sr I, Zr I, and Ba II are all strengthened. Lower: (c) Another spectral region of the *M*-type star, 56 Leonis; note that Tc I lines are weak or absent. (d) R Andromedae; note the strong lines of Tc I. The spectrum of R Andromedae was obtained by P. W. Merrill, and the upper two spectra by E. M. and G. R. Burbidge.

Crab Nebula



Supernova in IC 4182

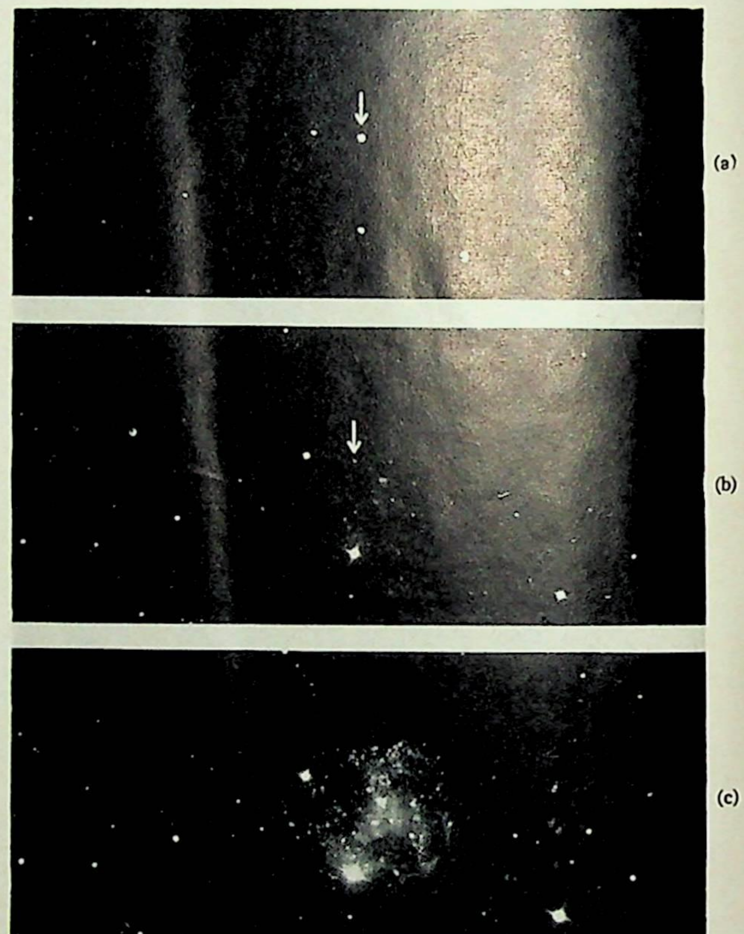


Plate 4. Left: The Crab Nebula, photographed in the wavelength range λ6300–λ6750. The filamentary structure stands out clearly at this wavelength, which comprises light mainly due to the Hα line. Right: The supernova in IC 4182, photographed (a) September 10, 1937 at maximum brightness—exposure 20 m; (b) November 24, 1938, about 400 days after maximum—exposure 45 m; (c) January 19, 1942, about 1600 days after maximum, when the supernova was too faint to be detected—exposure 85 m. Note that the lengths of the three exposures are different. These plates were taken by W. Baade, to whom we are indebted for permission to reproduce them.

Thus the calculated abundance is in fair agreement with that given by Suess and Urey. The expected decrease in cross sections with increasing N, neglected here, would give the observed gradual rise in abundance with N up to Hg^{202}. The drop in abundance at Hg^{204} is to be expected since it is not produced in the s process.

The large abundance of lead which we have derived from both the s and the r process results in considerable difficulties from the geological standpoint. Thus it is worthwhile to consider ways in which this conflict might be avoided. The most convincing evidence that the s process has been fully operative is the fact that the observed relative abundance of Pb^{208}, for which the r-process contribution is only 5%, and Pb^{204}, produced only by the s process, agree with the predictions of the s-process theory. Thus, the large Pb^{208}/Pb^{204} abundance ratio is attributable partly to the small neutron-capture cross section expected for magic Pb^{208} and partly to some cycling at the end of the s process. On the other hand, Pb^{204} might have an anomalously large cross section or might not be produced at all if Tl^{204} had such a large cross section that it captured a neutron and formed Tl^{205} instead of decaying to Pb^{204}. Neither of these possibilities seem likely, but they must be borne in mind as possible ways out of the conflict with geological evidence. It may be remarked that the total predicted lead production by the r process alone is 1.15 as compared with Suess and Urey's value of 0.47.

Finally, we consider the abundances of thorium and uranium. The recent results of Turkevich, Hamaguchi, and Reed (Tu56), using the neutron activation method, indicate that there is only a small amount of uranium in the Beddgelert chondrite. Urey (Ur56) has analyzed these results, which give an abundance of 0.007 on the scale of Suess and Urey. His results also give an atomic abundance of 0.02 for thorium. Our predicted r-process abundances are 0.147 for uranium and 0.462 for thorium. Thus, our calculations would seem to indicate that thorium and uranium as well as lead and bismuth have been reduced in abundance by some fractionation process in the formation of the planets and meteorites. As previously stated, of these four elements only lead has had its abundance in the sun determined (Go57), and in this case the solar abundance is in agreement with our calculated abundance and is very much higher than the abundance given by Suess and Urey.

In conclusion, the expected yield of radiogenic lead from the decay of thorium and uranium during the period since the formation of the meteorites would result in ratios of $(\mathrm{Rd}\,Pb^{206})/Pb^{204}$, $(\mathrm{Rd}\,Pb^{207})/Pb^{204}$, and $(\mathrm{Rd}\,Pb^{208})/Pb^{204}$ equal to 0.75, 0.41, and 0.57, respectively (see last line of Table VIII,4; Pb^{204} is taken as 0.2). These values are obtained by assuming no uranium-lead-thorium fractionation at formation. The Rd Pb^{206} and Rd Pb^{207} ratios are clearly consistent with Patterson's determinations of the radiogenic leads in chondrites and the Nuevo Laredo meteorite since we have used 4.5×10^9 years as the age of the meteorites. On the other hand, the ratios to Pb^{204} are considerably smaller than those found for radiogenic leads, indicating that lead was removed preferentially relative to uranium and thorium in chondrites and especially in the Nuevo Laredo meteorite. However, the iron meteorites could well contain the small amount of radiogenic leads which would be expected if no fractionation of lead relative to uranium and thorium occurred when they were formed.

IX. p PROCESS

In Sec. V it was pointed out that the proton-rich isotopes of a large number of heavy elements cannot be built by either the s or the r process, and particular examples of this difficulty are discussed in Sec. II, in connection with our assignments to one or the other of the synthesizing processes. With the exceptions of In^{113} and Sn^{115}, all of these isotopes have even A. From the Appendix we see that in the cases of Mo^{94}, Cd^{108}, and Gd^{152}, in addition to synthesis by the p process, these isotopes can also be made in weak loops of the s-process chain. In addition, In^{113}, Sn^{114}, and Sn^{115} may also be built in a weak loop of the s-process chain which is begun by decay from an isomeric state of Cd^{113}. In Fig. IX,1 a plot of the abundances of the p-process isotopes is given. The black dots represent isotopes which are made only by the p process, while the open circles represent isotopes which might have a contribution from the s process as well. A smooth curve has been drawn through the points merely to show the trends and positions of the peaks. This curve shows the same general trend as that of the main abundance curve, and the fact that the open circles lie predominately near the curve defined by the other points suggests that in no case is the s process an important contributor to the abundance. All of the isotopes are rare in comparison with the other isotopes of the same element, and it appears that only about 1% of the material has been processed by reactions which give rise to these isotopes (see Table II,1). The reactions which must be involved in synthesizing these isotopes are (p,γ) and possibly (γ,n) reactions on material which has already been synthesized by the s and the r processes.

The astrophysical circumstances in which these reactions can take place must be such that material of density $\gtrsim 10^2$ g/cm³ and containing a normal or excessive abundance of hydrogen is heated to temperatures of $2\text{–}3\times10^9$ degrees. It has been suggested (Bu56) that these conditions are reached in the envelope of a supernova of Type II. Alternatively they might be reached in the outermost parts of the envelope of a supernova of Type I in which the r process has taken place in the inner envelope. These possible situations are explored further in Sec. XII.

We suppose that for a short period quasi-statistical

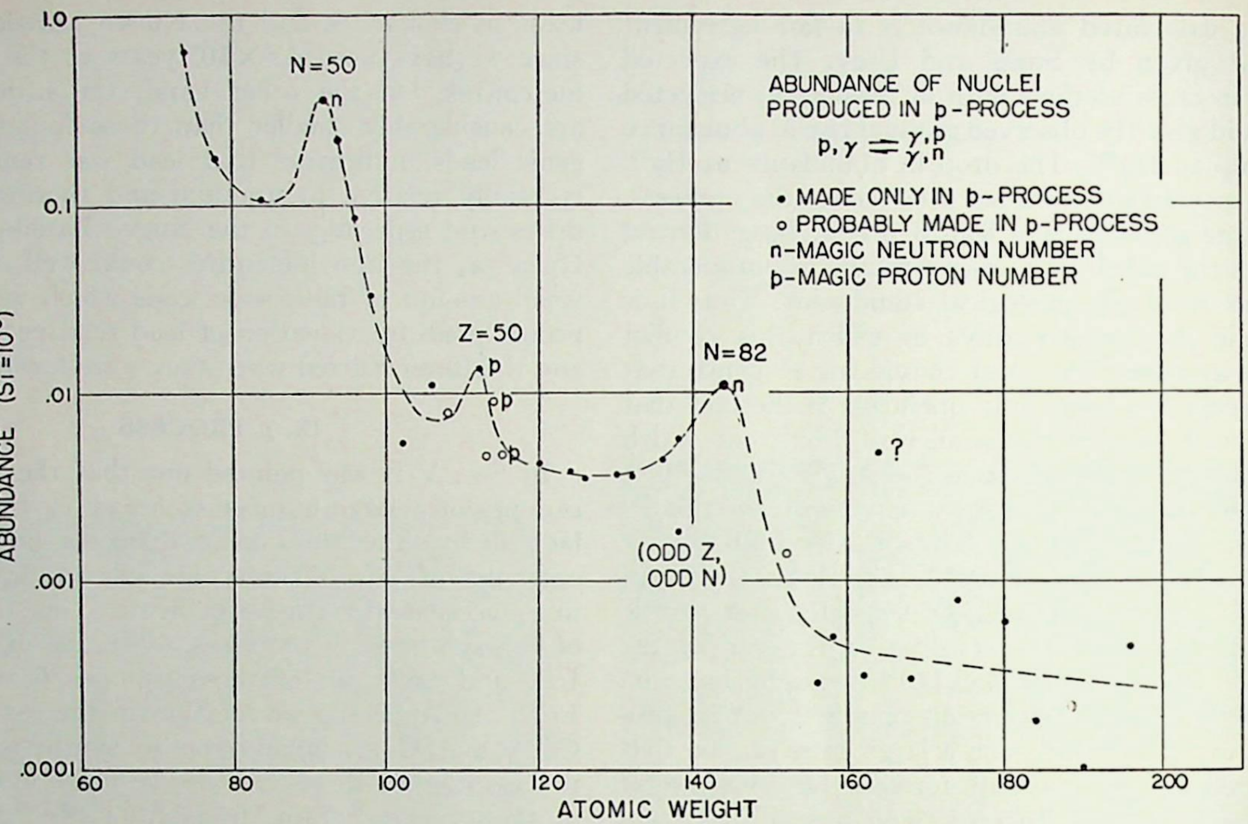


FIG. IX,1. Here we show a plot of the abundances of the isotopes made in the p process. The isotopes with magic N or magic Z are marked n and p, respectively. A curve has been drawn through the points to show the general trends. Note the peaks at $N=50$ and 82 and the lesser peak at $Z=50$.

equilibrium is reached between (p,γ), (γ,p), and (γ,n) reactions, i.e. $(p,\gamma)\begin{smallmatrix}\to(\gamma,p)\\ \to(\gamma,n)\end{smallmatrix}$. Since the initiating reaction is (p,γ), the flux of free neutrons built up by any (γ,n) reactions will not become comparable with the proton flux, so that complete equilibrium cannot be set up between protons, neutrons, and gamma radiation. In Fig. IX,2 we give a schematic diagram which shows what the effect of (p,γ) and (γ,n) reactions on nuclei which originally lie on the stability curve in the Z, A plane will be. These reactions tend to drive nuclei off the curve of greatest stability in the direction of increasing A and Z in the case of (p,γ) reactions and decreasing A in the case of (γ,n) reactions. Qualitative arguments suggest that the values of ΔA and ΔZ, the displacements off the main stability curve, will be small. The reasons for this are as follows. In the cases of ruthenium, cadmium, xenon, barium, cerium, and dysprosium, the two lightest isotopes built by the p process have roughly equal abundances or at least ratios which never exceed ~ 3. In the case of tin, three isotopes, Sn^{112}, Sn^{114}, and Sn^{115}, have abundances in the ratio $\sim 3:2:1$. Now if it were supposed that (γ,n) reactions on the heavier isotopes were mainly responsible for the production of those proton-rich isotopes we might expect, because of the cross-section effect, that in the case of tin, for example, the ratio would be more nearly like $1:x:x^2$ where $x \gg 1$. If, on the other hand, we supposed that (p,γ) reactions were responsible for driving the nuclei a very long way from the main stability line so that contributions from many nuclei were responsible for the tin isotopes, we would also expect that the ratios would not be small. Thus our qualitative conclusion is that (p,γ) reactions are responsible but the nuclei are only displaced in general two or three units of A and Z from the stability line. The existence of the nuclei Mo^{92} and Sm^{144} which have closed shells of 50 and 82 neutrons respectively and which show up as peaks in Fig. IX,1 suggests that these have been made directly from progenitors with closed neutron shells which form peaks in the normal abundance curve; i.e. in the case of Mo^{92}, these would be Zr^{90}, Y^{89}, Sr^{88}, etc., while the progenitors of Sm^{144} would be Nd^{142}, Pr^{141}, Ce^{140}, etc. The case of Er^{164} remains an anomaly, since the ratio $Er^{164}/Er^{162} \simeq 15$ is very large and is out of line with the results for the other pairs of isotopes. There appear to be no peculiarities in the possible progenitors of Er^{164} which might explain this large abundance.

A simple calculation can be made to see whether these qualitative results deduced from the observed abundances are correct. The equation of statistical equilibrium is

$$\log\frac{n(A+1, Z+1)}{n(A,Z)}=\log n_p-34.07-\tfrac{3}{2}\log T_9+\frac{5.04}{T_9}Q_p, \quad (26)$$

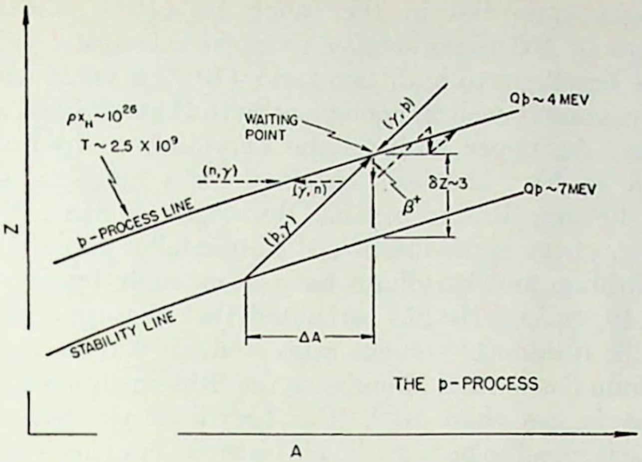


FIG. IX,2. The path of the p process in the Z, A plane. Material on the stability line (produced previously by the r or s process) is subjected to a hydrogen density of 10^{26} protons/cm^3 and a temperature of $\sim 2.5\times 10^9$ degrees. The (p,γ) reactions give an increase $\Delta A=\Delta Z=4$ or 5 until stopped by the (γ,p) reaction at nuclei with a proton binding energy, $Q_p\sim 4$ Mev. The (γ,n) reactions also produce a displacement onto the p-process line. Along this line positron emission must occur before further synthesis can take place. In general the lifetime of this emission is $\sim 10^3$ sec which is longer than the p-process conditions hold (100 sec). Thus a displacement $\delta Z\sim 3$ off the stability line (slope $\sim\frac{1}{3}$) at a given A occurs. This displacement by proton capture will occur even at large $A\sim 200$ in a time $\lesssim 1$ sec which is short compared to the time for the p process. Thus the displacement is terminated by the positron emission and not by decreasing Coulomb barrier penetrability so that it will be substantially independent of A. The p-process abundances as shown in Fig. I,1 and Fig. IX,1 should parallel the main abundance curve. The abundances will be 10^{-2} to 10^{-3} of those produced in the s and r processes.

where n_p is the number density of protons, Q_p is the proton binding energy, and T_9 is the temperature in units of 10^9 degrees. This equation is analogous to (14) in Sec. VII. Putting $n_p=10^{26}$/cm^3 and $T_9=2.5$, the binding energy of the last proton such that $n(A+1, Z+1)\approx n(A,Z)$ is obtained by putting the left-hand side of (26) equal to zero. We find $Q_p=4.3$ Mev. Now the proton binding energy is given by

$$Q_p(A,Z)=\alpha-\beta\left(1-\frac{4N^2}{Z^2}\right)-\tfrac{2}{3}\gamma A^{-\frac{1}{3}}-2\epsilon\frac{Z}{A^{\frac{1}{3}}}+\frac{\epsilon}{3}\frac{Z^2}{A^{\frac{4}{3}}}+g'(Z). \quad (27)$$

This equation is analogous to (23) of Sec. VII for the neutron binding energy; the symbols have the same meaning.

Thus

$$\left.\frac{\partial Q_p}{\partial A}\right|_N\approx -8\beta\frac{N^2}{A^3}\approx 0.6 \quad \text{for} \quad 100<A<200.$$

The binding energy of a proton in a nucleus on the main stability curve in this range of A is about 7 Mev. Thus $\Delta Q_p\simeq 2.7$ Mev and $\Delta A\simeq 4$ to 5. Thus ΔZ is also 4 to 5. However, we wish to know the deviation in Z of the new point from the stability line at $A+\Delta A$. We call this δZ and calculate it as follows. In Fig. IX,2 the slope of the line joining nuclei (A,Z) and $(A+1, Z+1)$ in the Z, A plane is 1. On the other hand, the change in Z corresponding to ΔA is just the change in the Z ordinate. Now the slope of the curve of maximum stability is $\sim\frac{1}{3}$. Thus δZ is given by

$$\delta Z=\tfrac{2}{3}\Delta A\simeq 3.$$

Whether the nuclei are driven off the main line to this maximum extent will depend on (i) whether sufficient protons are available, and (ii) whether the equilibrium conditions endure for sufficient time to allow the unstable nuclei to positron-decay so that the maximum number of protons can be added. The lifetime against positron emission is given by

$$\tau_{\beta^+}\simeq\frac{10^6}{W_{\beta^+}{}^5} \quad \text{(forbidden transition)},$$

where

$$W_{\beta^+}=B_A(\delta Z-0.5)+0.5.$$

Now $B_A\simeq 1.5$ Mev near $A=100$. The coefficient of B_A in this equation is $(\delta Z-0.5)$ instead of $(\delta Z-2.5)$ (which is used in Sec. VII) because we assume that the positron emission takes place by a forbidden transition to the ground state instead of by an allowed transition to an excited state as is the case for beta decay discussed in Sec. VII. In that case a mean value $\bar{W}_\beta$ was calculated.

Thus $W_\beta\simeq 4.2$, and $\tau_{\beta^+}\simeq 1000$ sec. Now the duration of a supernova outburst has been estimated to be 10–100 sec, and the time during which the p process takes place may be of the same order as or shorter than this explosion time. Consequently, it appears that the number of protons which can be added to the heavy nuclei is limited by the positron decay times and the synthesis to the limit of proton stability at this temperature will not be reached. Thus we conclude that the qualitative argument based on the observed abundances, that $\Delta A\sim 2$ or 3 and $\Delta Z\sim 1$ or 2, is borne out by this calculation.

The number of protons available to be captured by the heavy elements is determined by the number which are captured by the light and abundant elements. A typical example is C^{12}. Addition of two protons produces O^{14} whose half-life for positron emission is 72 sec. Thus it is probable that the maximum number of protons which can be added to C^{12} through the duration of the p process is only two. If we suppose that in the envelope in which the process occurs the H/C^{12} ratio is normal and equal to $\sim 10^4$, it is clear that even when we take into account all of the light elements which capture protons they can take only a small fraction of the total available, so that proton capture among the heavy nuclei is not limited by the number of protons available.

The mean reaction time for a (p,γ) reaction on a heavy nucleus A_0, Z_0 is given approximately by [see

Sec. III A, case (ii)]

$$\frac{1}{\tau}=3.1\times10^{9}\rho\frac{Z_0^{5/6}}{A_0^{1/6}T_9^{\frac{2}{3}}}$$
$$\times\exp\left[1.26\{AZ_0(A_0^{\frac{1}{3}}+1)\}^{\frac{1}{2}}-4.25\left(\frac{Z_0^2A}{T_9}\right)^{\frac{1}{3}}\right]\text{ sec}^{-1},$$

where $A=A_0/(A_0+1)$. For example, for $_{80}Hg^{200}$, we have

$$\frac{1}{\tau}=3.0\times10^{23}\frac{\rho}{T_9^{\frac{2}{3}}}\exp(-78.8T_9^{-\frac{1}{3}})\text{ sec}^{-1}.$$

Thus for $T_9=2.5$ and $\rho=10^2$ g/cm^3, $1/\tau=10$ sec^{-1}, so that the capture rate is rapid and is not a limiting factor in the proton addition.

X. x PROCESS

We have given the name x process collectively to mechanisms which may synthesize deuterium, lithium, beryllium, and boron. Some discussion of the problems involved in the x process are discussed in this section.

A. Observational Evidence for the Presence of Deuterium, Lithium, Beryllium, and Boron in our Galaxy

From the appendix the number ratios of deuterium, Li^6, Li^7, Be^9, B^{10}, and B^{11} relative to hydrogen are 1.4×10^{-4}, 1.8×10^{-10}, 2.3×10^{-9}, 5×10^{-10}, 1.1×10^{-10}, and 4.9×10^{-10}, respectively. Thus deuterium is rare as compared with its neighbors hydrogen and helium in the atomic abundance table, but as far as the remainder of the elements are concerned is very abundant and comparable with iron. Lithium, beryllium, and boron are all extremely rare as compared with their neighbors helium, carbon, nitrogen, and oxygen, and are only about 100 times as abundant as the majority of the heavy elements (see Fig. I,1). It should be emphasized for these elements particularly that all of these values have been obtained from terrestrial and meteoritic data and thus they may not be at all representative of the cosmic abundances of these elements.

A number of attempts have been made to detect deuterium in the sun. The latest, by Kinman (Ki56), shows that the abundance ratio of deuterium to hydrogen in the atmosphere is less than 4×10^{-5}. Attempts have also been made to detect the radio spectral line at 327 Mc/sec due to neutral deuterium in interstellar gas, and the most recent results by Stanley and Price (St56) and Adgie and Hey (Ad57) lead to the conclusion that the deuterium to hydrogen ratio does not exceed 5×10^{-4}.

The abundances of lithium and beryllium in the solar atmosphere have been investigated by Greenstein and Richardson (Gr51) and Greenstein and Tandberg-Hanssen (Gr54c). They found that while the isotope ratio Li^6/Li^7 could well be normal, the lithium to hydrogen ratio lies in the range $1.6\times10^{-11}-7\times10^{-12}$ or about 100 times smaller than the meteoritic value. The beryllium to hydrogen ratio $=10^{-10}$, a value which is reasonably well in agreement with that of Suess and Urey. An upper limit to the beryllium to hydrogen ratio of 10^{-10} has been obtained for a magnetic star by Fowler, Burbidge, and Burbidge (Fo55a). Estimates of the upper limits to the interstellar abundances of lithium and beryllium have been made by Spitzer (Sp49, Sp55). He has estimated that the upper limit to the lithium to sodium ratio is about 0.1, so that if sodium has normal abundance the lithium to hydrogen ratio is less than 10^{-7}. The beryllium to hydrogen ratio is found to be $\leq10^{-11}$. This is an order of magnitude lower than that found in meteorites, and Spitzer has concluded that some beryllium may be locked up in interstellar grains.

Lithium, probably in variable amounts, has been detected through the presence of the Li I resonance doublet at λ 6707.8 in a wide variety of late-type stars. Certain carbon stars show this feature very strongly in their spectra (Mc40, Mc41, Mc44, Sa44, Fu56), while it is also present in S-stars (Ke56, Te56), and in M-type dwarfs and giants (Gr57). No relative abundances are yet available.

Boron is spectroscopically unobservable.

In the primary cosmic radiation the abundances of lithium, beryllium, and boron are comparable with those of carbon, nitrogen, and oxygen (Go54a, Ka54, No55, No57).

B. Nuclear Reactions Which Destroy Deuterium, Lithium, Beryllium, and Boron

First we outline the reactions which destroy these light elements in a hydrogen-burning zone. At temperatures commonly found in the interiors of main-sequence stars the reactions which destroy deuterium are $D^2(d,p)H^3(\beta^-)He^3$, $D^2(p,\gamma)He^3$, and $D^2(d,n)He^3$. In the same way the reactions which destroy lithium are $Li^6(p,\alpha)He^3$ and $Li^7(p,\alpha)He^4$. Beryllium is destroyed by $Be^9(p,d)Be^8\rightarrow2He^4$ and $D^2(p,\gamma)He^3$, or $Be^9(p,\alpha)Li^6(p,\alpha)He^3$. Boron is destroyed by $B^{10}(p,\alpha)Be^7(EC)$-$Li^7(p,\alpha)He^4$ and $B^{11}(p,\alpha)Be^8\rightarrow2He^4$. Thus the net result is always to convert these elements into helium through proton bombardment, and the rates of the reactions are such that in all conditions before a star evolves off the main sequence all of the deuterium, lithium, beryllium, and boron in the volume which contains the vast majority of the mass will be destroyed (Sa55).

C. Synthesis of Deuterium, Lithium, Beryllium, and Boron

The foregoing considerations make it appear probable that these elements have been synthesized in a low-temperature, low-density environment in the universe, or conceivably in a region in which hydrogen was

absent. The alternative is a situation in which the synthesis was followed extremely rapidly by an expansion of the material resulting in cooling and decreased density so that the reactions leading to destruction were avoided, or else by a rapid transfer of the synthesized material to a low-temperature, low-density environment. Recent work of Heller (He57) shows very clearly the conditions of density and temperature under which a deuterium to hydrogen ratio $\simeq 10^{-4}$ can be preserved. Now the regions in our Galaxy in which such conditions are present are (i) in stellar atmospheres and (ii) in gaseous nebulae.

(i) This possibility was explored (Fo55a, Bu57) in connection with nuclear reactions which may take place in the atmospheres of magnetic stars (see Sec. XI). Besides having large magnetic fields and thus large sources of magnetic energy which may be available for particle acceleration and hence nuclear activity, the magnetic stars have one other feature which makes the results of elements synthesis in their atmospheres observable. It seems probable that their atmospheres do not mix appreciably with the lower layers. Thus the light elements synthesized there will not be mixed into the interiors and destroyed. However, the abundance of deuterium produced following the release of neutrons by (p,n) reactions on the light elements in such an atmosphere can never exceed a deuterium to hydrogen ratio $\sim 10^{-2}$. The use of different nuclear cross sections than those employed in Fo55a might conceivably raise this upper limit to $\sim 10^{-1}$. Because of the rarity of the magnetic stars and because of the small mass of their atmospheres, ejection of such atmospheres into the interstellar gas will lead to an interstellar deuterium to hydrogen ratio $\leq 10^{-10}$. If a more widespread class of stars such as the red dwarfs are proposed as stars in which deuterium can be made in the same way, a maximum interstellar ratio of 10^{-6} might be achieved. Thus since the lower limit to the interstellar abundance ratio so far determined is 5×10^{-4}, this explanation might be satisfactory. However, it leaves the terrestrial abundance to be explained by circumstances peculiar to the solar system, e.g., through electromagnetic activity in an early stage of formation of the solar nebula, a situation which is rather unsatisfactory.

Production of lithium, beryllium, and boron in a stellar atmosphere can take place through spallation reactions on abundant elements such as carbon, nitrogen, oxygen, and iron. Thus, if we believe that stellar atmospheres are the places of origin of these elements, it is also probable that they are a major source of the primary cosmic radiation, a conclusion which is consistent with observed abundances of primary nuclei mentioned earlier. Since energies $\gtrsim 100$ Mev/nucleon are demanded for spallation reactions it has been suggested that these reactions take place in regions fairly high in the stellar atmospheres. If it is supposed that these reactions go on only in magnetic stars, and that all the material synthesized is ejected, then the interstellar ratios of lithium, beryllium, and boron to hydrogen which can be obtained are all about 10^{-18}. However, if the red dwarfs are proposed as stars in whose atmospheres spallation reactions can go on at the same rates, ratios $\sim 10^{-10}$ can reasonably be predicted for the interstellar gas. Lithium has been found to be quite strong in the spectrum of T Tauri (He57a).

Thus we conclude that if the x process takes place in stellar atmospheres, a further process is demanded to synthesize the deuterium in the solar system, though the interstellar abundances of lithium, beryllium, and boron might be produced.

(ii) We shall now consider in a qualitative fashion the possibility that the x process takes place at some point in the evolution of gaseous nebulae. The current sources of energy in the bright gaseous nebulae are in general the highly luminous stars which are embedded in them. However, the discovery in recent years that some gaseous nebulae (not necessarily optically bright) are strong sources of nonthermal radio emission has shown that another source of energy must be present. There is strong support for the theory that this radiation is synchrotron emission by high-energy electrons and positrons moving in magnetic fields in these sources. The reservoirs of energy which must be present in such sources in the form of *high-energy particles and magnetic flux* are as large as the total amounts of energy which are released in supernova outbursts. Thus the energetic conditions that are demanded to produce a flux of free neutrons, which can then be captured to form deuterium and can also spallate nuclei to form lithium, beryllium, and boron, are also present. Unfortunately, the mean densities in such radio sources are so low ($\sim 10^{-21}$ g/cc) that at the present time the amount of activity leading to synthesis must be negligible. On the other hand, the situation in the early life-history of such nebulae may have been more conducive to such synthesis.

Recently the light from another star of the T Tauri class, NX Monocerotis, has been found to be largely polarized, thus indicating the presence of a considerable amount of synchrotron radiation (Hu57c). High-energy particles and a magnetic field must therefore be present. Stars in the T Tauri class are thought to be young stars, recently formed and not yet stabilized on the main sequence, and are embedded in dense interstellar matter. The observation of lithium in T Tauri itself and synchrotron radiation in NX Monocerotis support the idea that the x process (by spallation) is occurring here.

Of the strong radio emitters in our own galaxy, the Crab is known to be a supernova remnant, and it may be that some of the other sources are supernova remnants as well. In the context of the theory described in this paper, supernovae are the sources both of the r process and the p process, while a supernova outburst may well follow the e-process production of elements. Now if, in the early stages of a supernova outburst, after these other processes have taken place and the

envelope is beginning to expand so that the densities fall below those in which deuterium, for example. will be destroyed, the envelope can be subjected to a large flux of neutrons, it is possible that the x process could take place. However, as pointed out by Heller, if the source of neutrons which is demanded to synthesize deuterium is $C^{13}(\alpha,n)O^{16}$ or $Ne^{21}(\alpha,n)Mg^{24}$ (in Sec. III it was pointed out that there are strong reasons for rejecting $C^{13}(\alpha,n)O^{16}$ and using $Ne^{21}(\alpha,n)Mg^{24}$ as the source of neutrons here), then a high temperature $\sim10^9$ degrees is demanded to produce sufficient neutrons, but a low density $\sim10^{-4}-10^{-5}$ g/cm³ is demanded to preserve the deuterium; this appears to be an improbable situation. Thus a more plausible mode of synthesis would obtain if a large flux of neutrons from a high-energy, high-temperature region could be injected into an envelope of cool hydrogen at an early stage of the expansion. The following situation may be envisaged. As discussed in Sec. XII, at temperatures in excess of 5×10^9 degrees and densities $\geq10^8$ g/cm³ the stellar core will reach a configuration in which there can easily be a transition from a core made of iron to a core consisting primarily of helium with a fraction of free neutrons. If some portion of this core can be ejected into a surrounding, relatively cool, shell, the neutrons will be moderated and captured to form deuterium. The condition that they are captured before decaying is that $\bar{\rho}\geq10^{-10}$ g/cm³.

The very advanced stage of evolution in which a star might possess a neutron core might also provide a suitable source of neutrons for deuterium production, though this question will not be explored further here.

In Sec. XII it is shown that one percent of a supernova shell may be converted to the heavy elements in the r process. If the total mass of material is taken to be $\sim1M_\odot$, then dilution by a factor of the order of 10^4 is demanded to produce the correct abundances, relative to hydrogen, in the solar system, which is $\sim10^{-6}$ from Table II,1. Thus, the supernova shell must have mixed with a mass of about 10^4 times its mass in the interstellar gas. The deuterium to hydrogen ratio in the solar system is $\sim10^{-4}$ so it appears that if sufficient neutrons were available to convert the whole of the hydrogen in a supernova shell to deuterium, then this amount of dilution could explain the observed solar system abundance of this isotope. To produce the lithium, beryllium, and boron, it must be assumed that about 1% of the carbon, nitrogen, and oxygen in the shell was spallated either by high-energy neutrons or alpha particles ejected by the core or by protons into which the neutrons decayed. It is also possible that these latter elements may have been produced by a high-energy tail of the protons which are demanded in the p process. It does not appear probable, however, that a single supernova could account both for the deuterium and the r process elements in the solar system.

A supernova origin for the x-process elements would also suggest that supernovae are the original sources of cosmic radiation. Thus the primary cosmic radiation might be expected to have greater than normal abundances of iron and other abundant elements synthesized in supernovae. There is some evidence in support of this (No57).

D. Preservation of Lithium in Stars

Finally, we discuss an observational result which may suggest yet another region where lithium may be synthesized, and where it is certainly preserved. We summarized in Sec. X A the evidence that lithium is present in many kinds of late-type stars. Though no relative abundances of this element in different classes of late-type stars have yet been determined, it seems highly probable that the lithium abundance is variable, and also in some stars it is probably far more abundant than it is in the interstellar gas. This suggests strongly that there is some mechanism by which it can be preserved and even synthesized in stellar interiors. For example, in one S-type star, R Andromedae (Me56), both technetium and lithium are observed. Hence in this case there must be mixing between the helium-burning core where the technetium is synthesized and the surface. We have emphasized in Sec. X B that lithium cannot survive in a hydrogen-burning zone. Cameron (Ca55) attempted to show that Li^7 can be built in a hydrogen-burning zone through $He^3(\alpha,\gamma)Be^7(K)Li^7$, the Li^7 reaching the surface of the star because it is preserved in the form of Be^7, whose half-life against K capture is lengthened because of the comparative rarity of s-state electrons at the densities of production. Unfortunately his calculation of the rate of the reaction $He^3(\alpha,\gamma)Be^7$ is based on an estimate given by Salpeter (Sa52a) which was incorrectly printed, and it seems also that the increase of the half-life against K capture has been considerably overestimated.

Lithium does not react as rapidly with helium as it does with hydrogen, and it appears possible that it can exist in a helium-burning zone long enough for it to be able to reach the surface from the deep interior. Thus we conclude that the presence of lithium in the atmospheres of late-type stars suggests that these stars can have no hydrogen-burning zones. This implies that the helium cores must be large enough, and the stars' evolution sufficiently advanced, so that the hydrogen envelopes extend only to depths where the temperatures are less than $\sim10^6$ degrees. Reactions which may synthesize lithium in a helium-burning core will be discussed elsewhere.

XI. VARIATIONS IN CHEMICAL COMPOSITION AMONG STARS, AND THEIR BEARING ON THE VARIOUS SYNTHESIZING PROCESSES

Different processes of element synthesis take place at different epochs in the life-history of a star. Thus

the problem of element synthesis is closely allied to the problem of stellar evolution. In the last few years work from both the observational and theoretical sides has led to a considerable advance in our knowledge of stellar evolution. Theoretical work by Hoyle and Schwarzschild (Ho55) has been repeatedly mentioned since it affords estimates of temperatures in the helium core and in the hydrogen-burning shell of a star in the red-giant stage. However, the calculated model is for a star of mass 1.2 solar masses and a low metal content ($\sim$1/20 of the abundances given by Suess and Urey), and is thus intended to apply to Population II stars.

From the observational side, the attack has been through photometric observations of clusters of stars, from which their luminosities and surface temperatures have been plotted in color-magnitude or Hertzsprung-Russell (HR) diagrams, by many workers. We refer here only to two studies, by Johnson (Jo54) and by Sandage (Sa57a). A cluster may be assumed to consist of stars of approximately the same age and initial composition, and hence its HR diagram represents a "snapshot" of the stage to which evolution has carried its more massive members, once they have started to evolve fairly rapidly off the main sequence after their helium cores have grown to contain 10 to 30% of the stellar mass. Clusters of different ages will have main sequences extending upwards to stars of different luminosity, and the point at which its main sequence ends can be used to date a cluster; the rate of use of nuclear fuel, given by the luminosity, depends on the mass raised to a fairly high power (3 or 4).

Figure XI,1 is due to Sandage (Sa57a), and is a composite HR diagram of a number of galactic (Population I) clusters together with one globular cluster, M3 (Population II). The right-hand ordinate gives the ages corresponding to main sequences extending upwards to given luminosities. This diagram gives an idea of the way in which stars of different masses (which determine their position on the main sequence) evolve into the red-giant region. The most massive stars evolve into red supergiants (e.g., the cluster h and χ Persei); the difference between the red giants belonging to the Population I cluster M67 and the Population II cluster M3 should also be noted. Since stars evolve quite rapidly, compared with the time they spend on the main sequence, the observed HR diagram may be taken as nearly representing the actual evolutionary tracks in the luminosity-surface-temperature plane.

A feature of Population II stellar systems is that their HR diagrams contain a "horizontal branch" [see, for example, the work by Arp, Baum, and Sandage (Ar52)] which probably represents their evolutionary path subsequent to the red-giant stage. This feature is not included in the diagram of M3 in Fig. XI,1, although it is well represented by that cluster, since this diagram is intended to show just the red-giant branches of clusters. The old Population I cluster M67 has a sparse distribution of stars which may die on the Population I

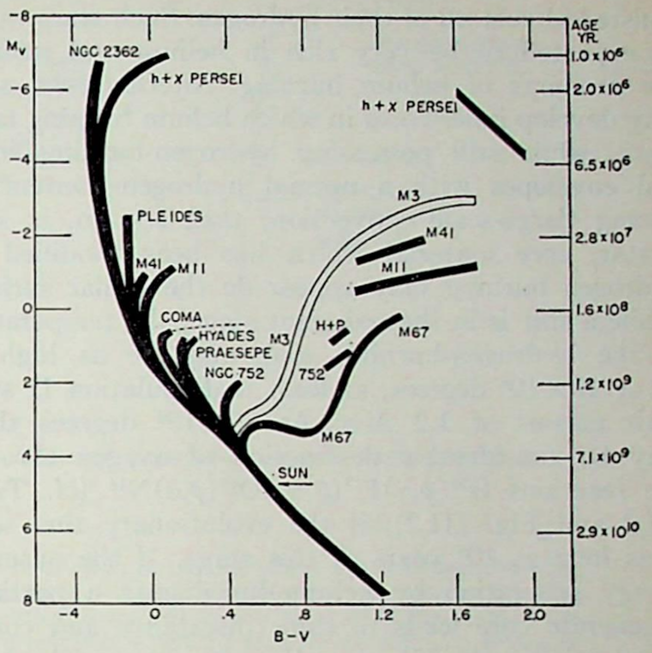


FIG. XI,1. Composite Hertzsprung-Russell diagram of a number of galactic (Population I) star clusters, together with one globular cluster, M3 (Population II), by Sandage (Sa57a). The abscissa measures the color on the B-V system, and defines surface temperature (increasing from right to left). The left-hand ordinate gives the absolute visual magnitude, M_v, of the stars (thus luminosity increases upwards). The heavy black bands (Population I) and unfilled band (Population II) represent the regions in the temperature-luminosity plane which are occupied by stars. The names of each cluster are shown alongside the appropriate band. The right-hand ordinate gives the ages of the clusters, corresponding to main sequences extending upwards to given luminosities. Note that the clusters all have a common main sequence below about $M_v = +3.5$. Note also that the red giants have different luminosities, according to the luminosities they had while on the main sequence (which are defined by their masses). *Diagram reproduced by courtesy of the Astrophysical Journal.*

analogy of the horizontal branch; the other Population I clusters in Fig. XI,1 do not show such a feature. Presumably the evolutionary history of a more massive Population I star subsequent to its existence as a red giant is more rapid.

Whenever reference is made in different parts of this paper to particular epochs in a star's evolutionary life, we are referring to a schematic evolutionary diagram for the star which has the same general characteristics as the HR diagrams in Fig. XI,1. With this background in mind, we turn now to astrophysical observations which provide many indications of element synthesis in stars. This is either taking place at the present time or else it has occurred over a time-scale spanned by the ages of nearby Population II stars. These indications may be divided into the following groups.

A. Hydrogen Burning and Helium Burning

It appears theoretically probable that some stars, perhaps delineated by their initial mass and amount of mass loss during their existence on the red-giant branch of their evolutionary path, may still exist as stable configurations for a time after they have ex-

hausted almost all of their hydrogen. Such stars would be expected to be very rich in helium and perhaps the products of helium burning. Alternatively, stars may develop inner cores in which helium burning takes place, while still possessing hydrogen-burning shells and envelopes with a normal hydrogen content. If mixing (large-scale convection) then sets in, in such a star, core material which has been modified by hydrogen burning may appear on the stellar surface. While a star is in the red-giant stage, the temperature of the hydrogen-burning zone may be as high as 30 or 40×10^6 degrees, at least in Population II stars with masses of 1.2 $M_\odot$. At 30×10^6 degrees there may be considerable destruction of oxygen through the reactions $O^{16}(p,\gamma)F^{17}(\beta^+\nu_+)O^{17}(p,\alpha)N^{14}$ (cf. Table III,2 and Fig. III,2), if the evolutionary time-scale is as long as 10^5 years at this stage. If the onset of energy generation by helium burning in a partially degenerate core leads to some instability and consequent mixing (Ho55), then the star may settle down into a new structure in which hydrogen burning occurs at a lower temperature, in material which may or may not have been appreciably enriched in C^{12}.

As was pointed out in Sec. III F(1), any stars in which the products of helium burning mix through a hydrogen-burning zone to the surface should appear rich in nitrogen and neon, as well as in helium, and considerably depleted in hydrogen. The equilibrium abundances of C^{13}, N^{14}, and N^{15} relative to C^{12} will be approached if sufficient time is spent in the hydrogen zone for the CN reactions to have gone through several cycles. On the basis of the most recent cross-section values N^{14} will be by far the most abundant of the carbon and nitrogen isotopes if equilibrium is attained even though C^{12} is the nucleus produced in the helium burning. O^{16} will also be converted to N^{14} (without cycling) if the temperature is great enough. Ne^{20} will mainly be converted to Ne^{21} and Ne^{22} and a very small amount of Na so that the element neon survives hydrogen-burning.

Greenstein (Gr54a), using the older rates of the CN cycle, discussed the connection between the operation of the CN cycle at various temperatures, the amount of mixing, and the abundance of nitrogen observed on the surface. We wish to elaborate on this argument at this point. If equilibrium is attained the ratio N/C will be given by 0.82 N^{14}/C^{12} since N^{15} is very rare and $C^{12}/C^{13}=4.6$ by number independently of temperature. The rough temperature dependence given for N^{14}/C^{12} in Sec. III F(1) is not sufficiently accurate for our present purposes and in Table XI,1 we give various equilibrium ratios as a function of temperature. In the last row of the table we list the logarithm of the time t in years for the N/C ratio to approach the equilibrium value at a given temperature from a not too greatly different value at another temperature. This time is of the order of magnitude of the sum of the mean lifetimes of C^{12} and C^{13} which is 22% greater than that of C^{12} alone and is thus easily calculable from the entries in Table III,2. In Table XI,1 we have taken $\rho x_H=10$ g/cc since this value is representative of the density and hydrogen concentrations in hydrogen-burning zones in giant stars. These characteristic times required to reach the equilibrium ratio for N/C are relevant to our problem. Only if the CN isotopes remain at the temperatures listed in the first row of Table XI,1 for times comparable to those listed in the last row will the N/C ratios given in the intervening rows be reached. The last and thus lowest temperature at which the CN isotopes spend sufficient time in mixing to the surface will determine the observed surface ratios. This assumes that the mixed-in CN outweighs that in the original envelope material. An inspection of the table indicates that this last, lowest temperature is probably about 15×10^6 degrees since the time required for N/C to reach its equilibrium value is $\sim10^7$ years at this temperature. We estimate that 10^7 years is the maximum time available for transport of the original C^{12} from the core through a significant portion of the envelope.

Figure XI,2 gives a schematic representation of the way in which the N/C ratio varies in the hydrogen-burning regions of a star of approximately four solar masses. During the initial condensation of the star, this ratio might be about 1 or 2, as in the solar system material and in young stars like 10 Lacertae and τ Scorpii (see Table XI,2). As hydrogen burning becomes established at central temperatures $20\text{–}30\times10^6$ degrees, the ratio rapidly reaches an equilibrium value in the range 100–50. When the star leaves the main sequence, gravitational contraction of the core occurs, the temperature of the hydrogen-burning shell rises and N/C decreases. At the onset of helium-burning, C^{12} is produced in the core and eventually some of this new carbon will be mixed into the hydrogen-burning regions. There will be a relatively short period during which the N/C ratio decreases to a minimum, followed by a rapid rise as N/C reaches its equilibrium value at the temperature at which the hydrogen burns in the star, which by now is in a highly evolved stage, per-

TABLE XI,1. Equilibrium nitrogen-carbon ratios as a function of temperature ($\rho x_H=10$).

$T(10^6$ degrees)	10	12	15	20	30	40	50	70	100
N^{14}/C^{12} by number	470	250	200	110	58	38	28	17	14
N/C by number	380	210	160	90	47	31	23	14	12
N/C by mass	440	240	180	100	54	36	26	16	13
$\log_{10}t$ (years)	10.6	9.0	7.2	5.1	2.4	0.8	−0.4	−2.1	−3.6

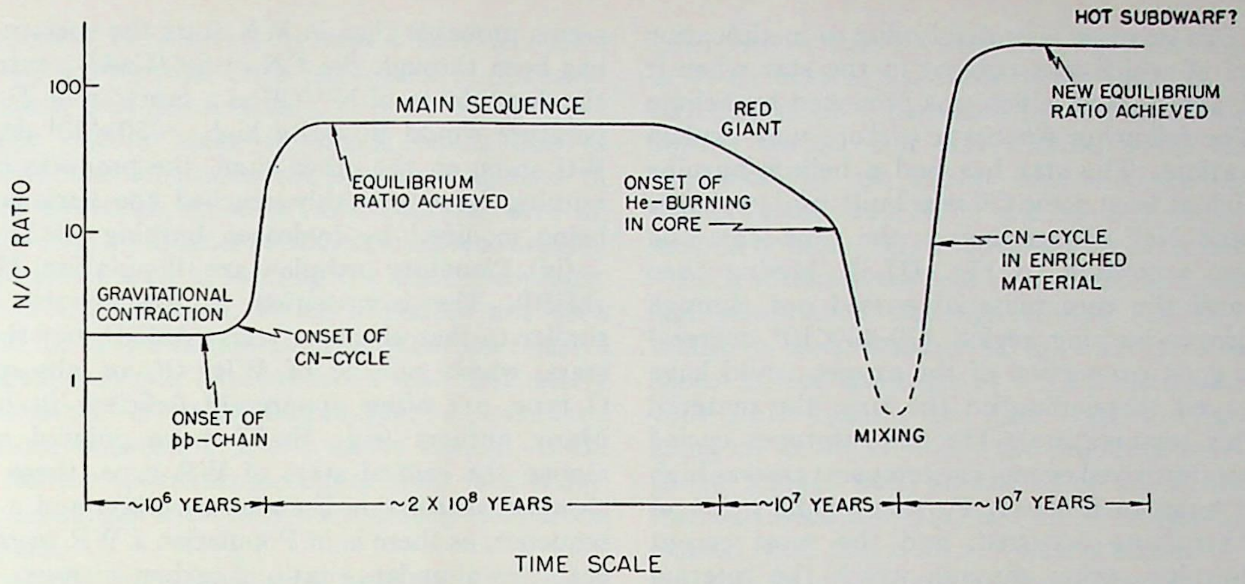


FIG. XI,2. Schematic representation of the variation of the N/C ratio in regions where the hydrogen burning occurs in a star of approximately four solar masses. An initial ratio of ~ 2, as in the solar system, is assumed. The time scales for different phases of the star's evolution are schematically marked as abscissas. With the onset of the CN cycle, an equilibrium ratio in the range ~ 50–100 is rapidly reached; when the star leaves the main sequence the temperature increases and the ratio drops. The onset of helium burning in the core reduces the ratio to an unknown lower limit defined by the onset of mixing. Passage of the core C^{12} through an outer hydrogen-burning region increases the ratio again to a new equilibrium value.

haps that of a hot subdwarf. The hydrogen-burning zone may be far out from the star's center and thus at a fairly low temperature, say 15×10^6 degrees. In this case we would estimate N/C to be ~ 160 with an upper limit from uncertainties in cross sections of ~ 250.

Several groups of stars exist whose spectra give evidence that hydrogen burning and helium burning have been occurring under the various possible conditions that we have just considered. These are as follows:

(i) Some stars, classified by other criteria as being of spectral types *O* or *B*, have no hydrogen lines; the stars HD 124448 (Po47), HD 160641 (Bi52), and HD 168476 (Th54) are examples. Helium and carbon lines are strong in all three; the oxygen lines usually prominent at this temperature are completely lacking in HD 168476. A preliminary determination of abundances in HD 160641 (Al54) has shown that carbon, nitrogen, and neon are all more abundant, relative to oxygen, than in normal stars, while helium has completely replaced hydrogen. These three stars have high velocities and may have originated in another part of the Galaxy than the solar neighborhood; they are probably Population II objects. The comparative rarity of such stars among surveys of *B*-type stars seems to indicate that they do not spend long in this evolutionary stage, or else the conditions under which stars evolve in this way are actually rare.

(ii) Münch and Greenstein have studied seven hot subdwarf stars with peculiar abundances of the light elements, and especially with an apparent excess of nitrogen. An analysis by Münch (Mu57a) of one such star, HZ 44, which is a high-temperature subdwarf, has shown that helium is more abundant than hydrogen. Nitrogen is about 200–300 times as abundant as carbon. Two determinations of the relative abundances of the elements, by mass, are given in Table XI,2, corresponding to the range of temperature and pressure given by the analysis. Aller's results for HD 160641 [see (i) above], are also given, together with the "normal" abundances by Suess and Urey, the abundances in the main-sequence (unevolved) young stars τ Scorpii and 10 Lacertae obtained by Traving (Tr55, Tr57), and the abundances in τ Scorpii obtained by Aller, Elste, and Jugaku (Al57c) (we have arbitrarily assumed here that the H/He ratio is the same as that determined by Traving). Since numbers of atoms are not conserved during element synthesis ($4H^1 \to He^4$, $3He^4 \to C^{12}$ etc.), only percentage abundances by mass are meaningful in considering the amount of increase or decrease in a given element, as compared with its normal abundance. In HZ 44 there appears to have been a slight increase in the sum of the elements heavier than helium, most of which is due to an increase in nitrogen by a factor between 10 and 25. The earlier discussion indi-

TABLE XI,2. A comparison between the chemical compositions of evolved stars and young stars.

Element	Percentage mass: Suess and Urey (Su56)	Main sequence young stars: Traving	Main sequence young stars: Aller	HZ 44 (Range)	HD 160641
H	75.5	58.6	[58.7]	9.0 – 9.1	0
He	23.3	39.8	[39.8]	88.7 –83.8	95.4
C	0.080	0.17	0.03	0.0047– 0.014	0.31
N	0.17	0.19	0.15	1.66 – 4.85	0.47
O	0.65	0.55	0.40	0.12 – 0.34	0.75
Ne	0.32	0.62	0.84	0.47 – 1.38	3.00
Si	0.053	0.10	0.07	0.17 – 0.48	0.065

cates that this increase is probably due to modification of C^{12}, part of which was present in the star when it condensed, and part of which was produced by helium burning. The following schematic picture may explain the observations. The star has had a helium-burning core in which at least some C^{12} was built, and probably some O^{16} and Ne^{20} (depending on the time-scale and temperature, according to Fig. III,3). Mixing then occurred, and the core material passed out through a hot hydrogen-burning region ($30–40\times10^6$ degrees) in which a good proportion of the oxygen could have been destroyed (depending on the time the material spent at this temperature). The neon isotopes cycled and were not destroyed even if the temperature was high enough for neon to interact. Further modification of the star's structure occurred, and the most recent hydrogen-burning region through which the internal material passed, on being mixed to the surface, was a cool one ($\sim15\times10^6$ degrees), in which the ratio N/C in the enriched material became $\sim$160 by number on the basis of our previous equilibrium estimates and $\sim$300 by observation. Since each of these values has errors of the order of $\pm50\%$ it would seem that they are in satisfactory agreement.

Münch has suggested that HZ 44 (and some other similar stars which he is studying at present) may be of Population I; they have low velocities and differ in other respects from "horizontal branch" stars of Population II, and may be analogous objects in Population I. The absolute magnitude of HZ 44 is in the range +3 to +5; the surface gravity seems high, and its mass may be considerably larger than 1 $M_\odot$.

(iii) The "classical" Wolf-Rayet stars, of both the carbon and the nitrogen groups, belong to Population I. A discussion, which includes references to earlier work, is given by Aller (Al43). There is still disagreement on the extent to which the characteristic broad emission features in the spectra are due to ejection of material, but there seems no doubt that these stars have reached a late evolutionary stage (Sa53). One example is a member of the binary system V 444 Cygni, in which the other component is an early-type massive star. The mass of the WR component is apparently less than the main-sequence component, but is still large ($\sim$10 $M_\odot$); if the star has reached a later evolutionary stage then its mass may originally have been larger. The lifetime on the main sequence may have been only a few million years. The WR stars are apparently deficient in hydrogen, but upper limits to its abundance have not been set. Helium is apparently the main constituent. In the WC stars the ratios by number of helium: carbon: oxygen are 17:3:1: (Al57a), and nitrogen is not seen at all. In the WN stars the ratios by number of helium: carbon: nitrogen: are about 20:1/20:1 (Al43, Al57a). All these estimates are rough, because of difficulties in abundance determinations, arising through large departures from thermodynamic equilibrium, stratification, etc. However, it seems probable that in WN stars the surface material has been through the CN cycle (Ga43); according to the dependence of N^{14}/C^{12} as a function of T, the temperature would be quite high, $\sim50\times10^6$ degrees. In WC stars, on the other hand, the products of helium burning have probably reached the surface without being modified by hydrogen burning (Sa53, Gr54a).

(iv) Planetary nebulae are Population II objects (Mi50). The composition of the nebulae is quite similar to that of young stars (Al57d), but the central stars, which can be of WR, *Of*, or absorption-line *O* type, are often apparently deficient in hydrogen. Many authors (e.g., Sw52) have pointed out that, among the central stars of WR type, there is not a clear-cut distinction between a carbon and a nitrogen sequence, as there is in Population I WR stars. Apparently the abundance ratio of carbon/nitrogen can take a whole range of values: the balance between the onset of mixing and the extent and temperature of a hydrogen-burning zone would seem not to be so critical as in the case of the Population I WR stars. The masses of the central stars of planetary nebulae are about 1 $M_\odot$ (Al56), and their absolute magnitudes are about 0 or +1. Bidelman (Bi57) and Herbig (St57) have suggested that the hydrogen-poor carbon stars of the R Coronae Borealis class [see (viii) below] may be the ancestors of some planetary nebulae.

(v) The *A*-type stars υ Sagittarii (Gr40, Gr47) and HD 30353 (Bi50) have anomalously weak hydrogen lines and appear to be rather similar to, although slightly cooler than, HZ 44 [see (ii)]. While helium, nitrogen, and neon are all strong in υ Sagittarii, carbon is weak. In Plate 1 (see pages 611–614 for plates) spectra are shown of the region near the Balmer limit in this star and in the *A*-type supergiant η Leonis (which it resembles in temperature and pressure), obtained by Greenstein at the McDonald Observatory. The comparison between the two stars is striking, the remarkable absence of the Balmer discontinuity in υ Sagittarii, and the strength of helium lines while at the same time ionized metals are strong, are impossible to explain other than by an abnormal composition. The great strength of lines other than hydrogen has been explained by the reduced opacity (mainly due to hydrogen in a stellar atmosphere of this temperature), but possibly the temperature is somewhat lower than has been estimated. Both υ Sagittarii and HD 30353 are spectroscopic binaries; this might be significant in view of work by Struve and Huang on mass-loss from close binaries (St57a), since one possibility for their evolutionary history is that the stars might have lost most of their hydrogen-containing atmospheres in this way.

(vi) Some white dwarfs, i.e., stars near the end of their life, show strong lines of helium and no hydrogen. These stars must have degenerate cores and the only nuclear fuel remaining must be in the surface layers. In Plate 1 we show Greenstein's spectra of three white dwarfs. One has nothing but strong helium

lines; one with very broad hydrogen lines is shown for comparison, together with one which is discussed in subdivision B of this section.

(vii) The carbon stars, in which bands due to carbon molecules dominate the spectrum instead of bands due to oxides, as in *M* and *S* stars, have an abnormally high abundance of carbon relative to oxygen, which will all be used in forming the spectroscopically inaccessible CO molecule. It has therefore not been possible so far to determine the ratio of carbon relative to hydrogen. Bouigue (Bo54a) has found that nearly equal abundances of carbon and oxygen, i.e., a carbon/oxygen ratio between 3 and 6 times normal, could account for the spectroscopic appearance (see Bi54). The carbon/oxygen ratio may vary among the carbon stars (Ke41).

The majority of carbon stars have an apparently normal atmospheric abundance of hydrogen. They may have a higher than normal abundance of the heavy elements (Bi54, Gr54a) (cf. paragraph D (iii) of this section], but no abundance determinations are yet available; technetium has been observed (Me56a). Again, the majority of these stars show bands of $C^{12}C^{13}$ and $C^{13}C^{13}$, indicating that the C^{12}/C^{13} ratio is about 3 or 4 (Mc48), which is near the equilibrium ratio produced in the CN cycle. Thus these stars may be ones in which helium burning ($3He^4 \rightarrow C^{12}$) has been taking place in the core (together, probably, with some neutron production leading to the *s* process). Large-scale mixing may then have set in *before* all the hydrogen in the envelope had been exhausted; the C^{12} passed through a hydrogen-burning shell on its way out to the surface, and the equilibrium ratio of C^{13} was thus produced. As many authors have noted, nitrogen should be abundant in these stars. Two examples of this kind of star, X Cancri (an irregular variable) and HD 54243, will be seen in Plate 2. Such stars at a later stage in their evolutionary history might look like WN stars or the star studied by Münch, HZ 44. However, the increased carbon/oxygen ratio might possibly be due to destruction of oxygen in a sufficiently hot hydrogen-burning region.

(viii) Some rare specimens among the carbon stars have weak or absent CH bands and hydrogen lines, and have been suggested to be deficient in hydrogen (Wu41, Bi53, Bu53). There is no evidence to suggest that in these stars the C^{12}/C^{13} ratio is smaller than that in the solar system ($\sim$90). Most of the group are variables of the R CrB type; RCrB itself was analyzed by Berman (Be35) and found to be very rich in carbon and deficient in hydrogen. There are a few nonvariable examples, two of which are shown in Plate 2. A spectrum of HD 137613 (Sa40, Bi53) has been borrowed from the Mount Wilson files and is shown alongside the two normal carbon stars; absence of C^{13} bands and weakness of H_β are striking. The stars are not a good match in temperature, and hence differences in their line spectra will be seen; however, HD 137613 is hotter than the other two so that H_β would be expected to be stronger in it. Greenstein's spectra of the peculiar carbon star HD 182040, first discussed by Curtiss (Cu16), and the normal carbon star HD 156074, which have similar temperatures, are shown together; the weakness of the CH band and H_γ in the former, while C_2 can be seen, is notable. HD 182040 has no detectable C^{13}.

Such stars are too cool to show helium even if it is in high abundance, although Herbig (He49a) suggested tentative identification of an emission feature at λ 3888.4 in the spectrum of R CrB near minimum light with HeI. Possibly the $3He^4 \rightarrow C^{12}$ reaction has been going on in the core and the products were not mixed to the surface until almost all of the hydrogen in the star had been used up, and no hydrogen-burning shell existed, and hence no C^{13} was produced. The R CrB stars probably belong to Population II. Stars of this sort at a later evolutionary stage might look, for example, like HD 124448, or like those central stars of planetary nebulae which are of WC type, as discussed in (iv) above.

Two stars with a large C^{12}/C^{13} ratio, HD 76396 and HD 112869 (Mc48), have excessively strong CH; they belong to the high-velocity group of CH stars studied by Keenan (Ke42). Another star of the same sort is HD 201626 (No53). If C^{12} has actually mixed out to the surface, then it cannot have gone through a hydrogen-burning region. Nitrogen should then be less abundant than in normal carbon stars. One possibility is that these stars do not possess a hydrogen-burning shell, but have a relatively thin surface region in which hydrogen still exists. Another possibility is that the temperature of an outer hydrogen-burning zone was not high enough. The high-velocity (Population II) character of the CH stars means that they may, at the beginning of their lives, have had a ratio of carbon, nitrogen, and oxygen to hydrogen which was much lower than in Population I stars.

(ix) Miss Roman (Ro52) has found that certain *G*- and *K*-type giants having high velocities have a peculiar appearance in the cyanogen band; she has designated these the "4150" stars. Bidelman (Bi57) suggested that these stars may have a higher than normal carbon abundance. However, he has commented recently that, although these stars have strong cyanogen (CN), they do not show C_2 bands (private communication). Perhaps they are also examples of stars whose surface layers contain material that has recently passed through the CN cycle; the nitrogen/carbon ratio may be larger than normal.

A peculiar *G*-type giant, HD 18474 (Bi53, Hu 57a) and two similar stars studied by Greenstein and Keenan, have weak CH and somewhat weakened CN absorption; these seem to have a low carbon/hydrogen ratio, while the abundance of the metals seems to be normal in two of the stars and low in one which has a high velocity (Gr57a).

To summarize, we conclude that the spectroscopic

indications of the occurrence of hydrogen burning and helium burning in stars show great diversity. Hydrogen burning is demonstrated by complete hydrogen exhaustion (e.g., HD 160641, some white dwarfs); partial hydrogen exhaustion and high nitrogen abundance (e.g., HZ 44, WN stars, υ Sagittarii); occurrence of C^{13} in the equilibrium ratio with C^{12} (e.g., normal carbon stars). Knowledge of the masses of the various stars mentioned here is scanty, but a wide range may be involved, leading to a range in evolutionary tracks and time-scales (see Fig. XI,1). The general evolutionary sequence, however, is probably as follows: (a) the carbon stars (red giants); (b) stars like HD 160641, the WN stars, and HZ 44 ("horizontal branch" stars and their analogy in Population I); (c) the white dwarfs (exhaustion of fuel, end of life).

Evidence for helium burning is less direct. The main problem is that in cool stars the depletion of oxygen during hydrogen burning will always lead to an apparent increase of carbon, because less CO will be formed. Further study of dissociation equilibria of the various molecules, together with a knowledge of the opacity of carbon stars, may throw more light on the problem. However, the WC stars and possibly HD 124448 may have a large increase in carbon; in HZ 44 a smaller increase may have occurred (see Table XI, 2).

The way in which conditions for hydrogen and helium burning depend on the initial mass of the star is at present quite uncertain, and more computed evolutionary tracks, carried through the onset of helium burning, are needed.

B. α Process

A peculiar white dwarf, Ross 640 (Gr56), has strong features attributed to magnesium and calcium. Greenstein's spectrum of this star is reproduced in Plate 1. Although analyses of white dwarf atmospheres, with their extremely large surface gravities, have not been carried out, it seems hard to explain such a spectrum except by the actual presence of large amounts of the α-process elements magnesium and calcium.

One peculiar and perhaps unique white dwarf has no recognizable spectral features except very broad absorptions at λλ 3910, 4135, and 4470. Greenstein (Gr56a) has suggested that these may be the helium lines λλ 3889 +3965, 4121+4144, and 4438+4472, arising under conditions of extreme pressure, with the obliteration of atomic orbits which would give rise to other lines normally expected. Burbidge and Burbidge (Bu54) have suggested that the features at λλ 3910 and 4135 may be due to Si II λλ 3854–3863 and 4128–4131, and that the feature at λ 4470 may be due to Mg II 4481, again under very high pressure, and have proposed that the star has a high abundance of the α-process elements magnesium and silicon.

C. Synthesis of Elements in the Iron Peak of the Abundance Curve, and the Aging Effect as It Is Related to this and Other Types of Element Synthesis

Stars which have evolved so far as to have the central temperatures and densities necessary for the equilibrium production of elements in the iron peak become unstable (see Sec. XII). Thus if they spend only a short time in this stage, specimens in which it is occurring or has recently taken place would be expected to be rare (unless stellar remnants—white dwarfs—are left after the explosion). However, if element synthesis has been going on throughout the life-history of our galaxy we might expect to observe an aging effect in that the oldest stars might have the lowest metal abundances. The following observations support this.

(i) Some stars in whose spectra metallic lines are strikingly weak have been observed. Analysis by Chamberlain and Aller (Ch51) of the so-called subdwarfs, HD 19445 and HD 140283, led to an iron abundance 1/10 of normal and a calcium abundance 1/30 of normal for these stars. A recent more detailed analysis of HD 19445 by Aller and Greenstein (Al57b) has confirmed the low iron abundance, which they find to be 1/20 of normal. Plate 2 shows the photographic region of the spectrum of HD 19445 together with the normal *F*7 V star ξ Peg; HD 19445 actually has a slightly lower temperature than ξ Peg (Gr57), and if it were compared with a *G*0 or *G*2 star (where its temperature places it) the difference in the strength of the metallic lines would be even more striking than illustrated in Plate 2. The absolute magnitude of HD 19445, about +5, is very near to the main sequence at the true temperature of the star; certainly the departure from the main sequence is within the error in the trigonometric parallax. The name "subdwarf" is thus seen to be misleading. Most of the "subdwarfs" in this temperature range have been so designated from spectroscopic parallaxes, which are uncertain since the spectral peculiarities of these stars have caused them to be classified too early on low-dispersion spectrograms.

HD 19445 and HD 140283 have high space velocities and belong to Population II. The implication is that they are members of a spherical distribution of stars which condensed before the Galaxy assumed its present flattened form and hence are old stars. Similar extreme weak-line stars, presumably belonging to Population II, e.g., HD 122563 (Gr57), may have low velocities. The fact that calcium is weakened as well as the iron-peak elements means that this aging effect shows, as would be expected, in other element-building processes besides the *e* process. Plate 2 shows that a line of Sr II is also weak (strontium is produced by the *s* process).

Five more stars which have some of the characteris-

tics of Population II were analyzed by Burbidge and Burbidge (Bu56a), by the curve-of-growth method without allowing for the effect of a lowered metal abundance on the opacity. In each star the abundances of a selection among the elements magnesium, aluminum, calcium, scandium, titanium, chromium, manganese, iron, strontium, yttrium, zirconium, and barium were determined. The stars showed a spread in the ratio of the abundances of these elements relative to hydrogen which indicated a range in ages; the most extreme case had a similar ratio to those in HD 19445 and HD 140283. The abundance ratios for strontium, yttrium, and zirconium showed no significant differences from those for iron, but tentative results for barium gave smaller ratios than those for iron (this result was however, uncertain).

(ii) Stars with high space velocities do not belong to the solar neighborhood, and in particular those with high z velocities may be presumed to belong to an older more spherical population. The so-called subdwarfs have very high space motions, but Miss Roman (Ro50, Ro52) and Keenan and Keller (Ke53) found a general correlation between space motion and spectral peculiarities in that high-velocity stars tend to have strong CH and weak CN and metallic lines. A recent analysis of a group of high-velocity stars by Schwarzschild, Schwarzschild, Searle, and Meltzer (Sc57b) has shown that the most extreme example measured by them, ϕ^2 Orionis, has ratios of the metals to hydrogen and of carbon, nitrogen, and oxygen to hydrogen, that are both one quarter of the solar values. The heavy elements also are under-abundant by the same factor as iron. Thus if the different compositions of the so-called subdwarfs, the less extreme high-velocity stars, and the solar-neighborhood stars are due to an aging effect, then elements produced by hydrogen burning, helium burning and the α, e, and s processes are all affected, and in the high-velocity stars the factors are all about the same. Seven high-velocity G-type giants analyzed by Greenstein and Keenan (Gr57a) have values for the metals/hydrogen ratio varying from 0.4 to 0.6 times normal.

(iii) The globular clusters may contain the oldest stars in our galaxy. The age of M3 has been given as 6×10^9 years by Johnson and Sandage (Jo56). It might therefore be expected that the stars in such a cluster would have lower abundances of the metals and other elements than normal. Also, theoretical evolutionary tracks can be made to agree with the observed color-magnitude diagram only if a low metal abundance is used (Ho55). Unfortunately, the brightest main-sequence stars of this cluster (in which turbulence and opacity should have only small effects on abundance determinations) are too faint for spectrophotometric analysis. However, information may be derivable from the colors. The two-color (U−B, B−V) plot for M3 lies well above the normal solar-neighborhood plot throughout its length, while the so-called subdwarf HD 19445 [see C (i)] lies very near to the M3 plot. This position of HD 19445 may be the result of different "blanketing" of the continuous spectrum by the spectrum lines. All lines are weakened in HD 19445; the metallic lines are more numerous in the blue (B) than in the visual (V) spectral region, and still more numerous in the ultraviolet (U). The effect of weakened metallic lines would therefore be to increase the blue intensity relative to the visual, and to increase the ultraviolet relative to the blue. Qualitatively therefore, as discussed by Johnson and Sandage (Jo56), this is in the same sense as that required to explain the abnormal position of HD 19445 on the U−B, B−V diagram. An accurate evaluation of the effect is currently being made by Sandage, Burbidge, and Burbidge. If this explanation is found to be satisfactory then we shall have indirect proof that the stars of M3 have low metal abundances. Sandage has shown by an approximate calculation that if the amount of light subtracted from the sun's continuum by the spectrum lines is added in the appropriate regions (without allowing for any redistribution of energy caused by blanketing), the effect is to shift the sun on to the M3 line in the (U−B, B−V) plot, at a value of B−V corresponding to a normal $F5$ star.

(iv) Studies of colors and spectra of external galaxies are still too uncorrelated for us to make deductions concerning abundance variations. It may, however, be worth entertaining the possibility that some of the elliptical galaxies which have much earlier-type spectra than would correspond to their color classes may prove to have low calcium and iron abundances, relative to hydrogen.

(v) Apart from the foregoing discussion of possible aging effects (in other processes besides the e process), no undoubted examples of stars exhibiting the e-process abundances predominantly are known. However, a speculative possibility may be considered. The white dwarf van Maanen 2 shows strong, very broadened lines of iron (Gr56); possibly it has a high abundance of iron. Such stars may be the remnants (perhaps after large mass loss) of stars which reached a central temperature sufficient to build the iron peak, but did not suffer complete catastrophic explosion.

D. *s* Process

Three groups of giant stars have abundance anomalies that indicate the occurrence of the *s* process in their interiors. These are (i) the *S* stars; (ii) the Ba II stars; and (iii) the carbon stars.

(i) The *S* stars probably have luminosities in the range of normal giants. Elements with neutron magic-numbers appear more predominantly in the spectra of *S* stars than in *M* stars of the same temperature (Me47, Al51, Bu52, Bi53a, Me56, Te56). It has been realized for some time that this must be a real atmospheric abundance effect and not an ionization,

TABLE XI,3. Overabundance ratios for the sum of the *s*-process isotopes of each element in HD 46407, calculated on the assumption that the observed overabundance ratios are due solely to building of the *s*-process isotopes.

Element	Observed ratio y	Suess and Urey abundances s	$r+p$	Calculated ratio for *s*-isotopes y'	$\langle\sigma N\rangle y'$
Strontium	4.7	18.79	0.106	4.7	381
Yttrium	7.8	8.9	···	7.8	1326
Zirconium	4.8	52.92	1.53	4.9	884
Niobium	5.2	1.00	···	5.2	832
Molybdenum	3.2	1.4185	1.0015	4.8	343
Ruthenium	7.8	0.6445	0.8452	16.7	1837
Barium	14	3.4326	0.2253	14.9	281
Lanthanum	9.8	2.00	0.0018	9.8	647
Cerium	9.6	2.00	0.2601	10.7	407
Praseodymium	28	0.40	···	28	672
Neodymium	14	1.276	0.1630	15.7	443
Samarium	13	0.3080	0.3448	26.4	4050
Gadolinium	3.0	0.0147	0.668	93	3255
Ytterbium	2.0	0.12346	0.0944	2.8	78
Tungsten	9	0.326	0.1646	13.0	914

dissociation, or excitation effect. Fujita (Fu39, Fu40, Fu 41) suggested that the *S* stars have carbon abundances intermediate between *M* and the carbon stars. The appearance of strong lines of Tc I in *S* stars (Me52) has been previously described as a good indicator of current nuclear activity and mixing on a time scale $\sim 10^5$ years in these stars. In Plate 3 we show two portions of the spectrum of the *S*-type long-period variable star R Andromedae obtained by Merrill, together with a standard *M*5 III star; this spectrum was taken from the plate files at Mount Wilson Observatory. The first portion shows the lines of Tc I in R Andromedae; the second shows a region where ZrO bands are present in R Andromedae, and TiO bands are seen in the *M*-type star. The strength of Ba II λ 4554 in R Andromedae will also be noted. The two stars are not a perfect match in temperature, so that some differences in the line spectra will also be noted. R Andromedae is somewhat cooler than the *M*-type star, and since it is a long-period variable, emission lines will be seen in its spectrum, notably at H_γ.

(ii) From the spectroscopic point of view the rarer Ba II stars (Bi51) represent an easier problem to analyze, since their temperatures are considerably higher, their spectra being of types *G-K*, while their spectral peculiarities link them closely with the *S* stars. They have a greater-than-normal abundance of carbon, probably intermediate between the *M* and the carbon stars. A recent curve-of-growth analysis of HD 46407, a typical Ba II star (Bu57a), has yielded abundances of a large number of elements relative to those in the standard *G*8 III star κ Geminorum, which is very similar in temperature and luminosity and was used as a comparison. Portions of the spectra of HD 46407 and κ Geminorum are shown in Plate 3; the most prominent of the features that are strengthened in the Ba II star are marked. Abundances of all elements which

TABLE XI,4. Search for *r*-process elements in HD 46407.

Element	Whether observed in the sun	Sensitive lines: whether accessible	Whether strengthened in HD 46407
Selenium	No	No	No evidence
Bromine	No	No	No evidence
Krypton	No	No	No evidence
Tellurium	No	No	No evidence
Iodine	No	No	No evidence
Xenon	No	No	No evidence
Osmium	Yes	Yes	Inconclusive[a]
Iridium	Yes	Yes	Probably no[b]
Platinum	Yes	Yes	Very probably no[c]
Gold	Yes	Yes	No evidence[d]

[a] All lines in the observed wavelength range of HD 46407 are too badly blended for any identifications to be made.
[b] At the wavelengths of the strongest lines in the observed range, *no lines* are visible, either in HD 46407 or in the standard star.
[c] There are four possible identifications, for all of which the lines in HD 46407 are equal to those in the standard star.
[d] Sensitive line is outside the observed range in HD 46407. Other lines are not seen.

could be studied, lighter than $A=75$, and in particular elements in the iron peak, are normal. Most of the elements with $A>75$, which are spectroscopically observable, are overabundant; this includes in particular those with neutron magic-numbers. The observed abundance ratios, relative to normal, are given in column 2 of Table XI,3. With the possible exception of gadolinium, all these elements have a large proportion of their isotopes produced in the *s* process, as is shown in the appendix. On the other hand, with the possible exception of ytterbium, all of the elements with $A>75$ which we have assigned mainly to the *s* process, and which are spectroscopically observable, are found to be overabundant.

The observations in HD 46407 refer only to the sum of all isotopes of each element, and we wish to derive the true overabundances of those isotopes actually built by the *s* process. From the appendix we have obtained the sum of the solar-system abundances of the isotopes of each element in Table XI,3 which are built by the *s* process, and by the *r* and *p* processes together. If these sums are denoted by s and $r+p$, respectively, and if the observed and true overabundances are y and y', respectively, then we have

$$y(s+r+p)=y's+r+p.$$

Values of s, $r+p$, and y' are given in columns 3, 4, and 5 of Table XI,3.

Finally, we have taken the mean vale of σN for all the *s*-process isotopes of each element, and the product of this and y' is given in the last column. Under conditions of steady flow this product should be constant (see Sec. VI), and this is the case to within a factor of about 2, if we exclude ruthenium, samarium, gadolinium, ytterbium. Of these, only samarium has a well-determined abundance, and the possibility that cross sections for this element in the appendix are too large was discussed in Sec. VI. The estimates of σ for Gd^{154}, Gd^{155}, and Yb^{89} may also be too large, while

that for Ba^{134} may be too low. If further abundance determinations confirm the low value of y for ytterbium then it may be found that a smaller proportion of its abundance should be assigned to the s process.

Europium, both of whose isotopes have been assigned to the r process, turned out not to be overabundant in HD 46407. A search for the most abundant of the r process elements in the spectrum of HD 46407 was inconclusive; most of these elements cannot be identified in HD 46407 or in the standard star. However, none were found to be strengthened in HD 46407. The results of this search are given in Table XI,4. This table is given in order to show how difficult it is to identify spectrum lines due to many of the r-process heavy elements, let alone determine abundances, in stars other than the sun.

The elements copper, zinc, gallium, and germanium are all produced predominantly by the s process, yet the first two and probably the last are not overabundant in HD 46407 (we have no evidence concerning gallium). These elements fall on the steep part of σN *versus* A plot in Fig. VI,3, and this suggests, therefore, that there is a plentiful supply of neutrons produced in Ba II stars, giving the shorter of the two capture-time-scales discussed in Sec. II, and leading to building of elements according to the flat part of the curve in Fig. VI,3. If this process occurs in a volume containing $10^{-2}-10^{-3}$ of the mass of the star, then subsequent mixing could produce the observed overabundances; normal abundances would be expected to be observed for copper, zinc, gallium, and germanium, and depletion of the iron-peak elements would not be detectable. Such a copious flux of neutrons means that interior temperatures of $\sim 10^8$ degrees must have been reached. However, we have independent indications, from the presence of an excess amount of carbon in this and other Ba II stars (as also in the S stars), that a temperature sufficient for the reaction $3He^4 \rightarrow C^{12}$ to occur has been achieved.

Tc I lines have not been observed in Ba II stars; before it can certainly be said that technetium is absent a search should be made for Tc II in the ultraviolet. Possibly the Ba II stars represent a later evolutionary stage than the S stars, or they may evolve from stars in a different mass range in which mixing to the surface takes a time $\gg 10^5$ years.

HD 26 is a Ba II star with Population II characteristics (abnormally strong CH, weak metallic lines, and a high velocity). The strongest spectrum lines include those of the rare earths [Vrabec, quoted by Greenstein (Gr54a)]. This suggests that the s process is also taking place in the oldest stars.

(iii) As was mentioned in paragraph A(vii), the carbon stars also have apparent overabundances of the heavy elements, less striking than in the S stars, together with the presence of technetium. Actual abundances have not been determined, but it seems as though the s process has been occurring here also, perhaps in a smaller volume of the star, while a core containing more carbon is developed. More abundance determinations, together with computations of evolutionary tracks of stars, will be helpful in elucidating the evolutionary differences between M, C, S, and Ba II stars.

E. r Process

The outstanding piece of observational evidence that this takes place is given by the explanation of the light curves of supernovae of Type I as being due to the decay of Cf^{254} (Bu56, Ba56), together with some other isotopes produced in the r process. Further evidence can be obtained only by interpreting the spectra of Type I supernovae, a problem which has so far remained unsolved.

F. p Process

No astrophysical evidence is at present available for the occurrence of this process. The present theory might suggest that isotopes built in this way should be seen in the spectra of supernovae of Type II.

G. x Process

The situation as far as the astrophysical circumstances of the x process are concerned has been discussed in Sec. X.

H. Nuclear Reactions and Element Synthesis in the Surfaces of Stars

The peculiar A stars, in which large magnetic fields have recently been found (see Ba57), have long been known to have anomalously strong (and often variable) lines of certain elements in their spectra. Abundances of all the observable elements have been determined in two magnetic stars, α^2 Canum Venaticorum and HD 133029, and in one peculiar A star, HD 151199, in which a magnetic field has not been detected, owing to the greater intrinsic width of the spectrum lines, but may well be present (Bu55, Bu55a, Bu56b). The results are given in Table XI,5, which should be compared with Table XI,3 and the discussion of the Ba II star HD 46407. At first glance the results are very similar, particularly in the fact that predominantly the elements with neutron magic-numbers are overabundant, but there are some striking differences, as follows: (a) barium is overabundant in HD 46407, and normal in the peculiar A stars; (b) europium is the most prominently overabundant element in the peculiar A stars, and is normal in HD 46407; (c) silicon, calcium, chromium, and manganese are normal in HD 46507 while calcium is underabundant and the rest are overabundant in the peculiar A stars; and (d) the overabundance ratios of the strontium group and the rare earth group differ by factors of about 2 in HD 46407 and by factors varying from 2 in HD 151199 to 40 in α^2 Canum Venaticorum.

TABLE XI,5. Abundances of the elements in three "peculiar *A*" stars, relative to those in a normal star.

Element	Magic number isotope (*A*,*N*) for anomalous elements	α^2CVn	HD 133029	HD 151199
Magnesium		0.4	1.4	1.2
Aluminum		1.1	2.2	
Silicon	28, 14	10	25	1.3
Calcium	40, 20	0.02	0.05	2.6
Scandium		0.7		
Titanium		2.6	2.3	
Vanadium		1.3	3.2	
Chromium	52, 28	5.2	10.3	1.8
Manganese		16	15:	9
Iron		2.9	4.1	1.1
Nickel		3.0	2	
Strontium	88, 50	14	11.5	65
Yttrium	89, 50	20:		
Zirconium	90, 50	30	40	
Barium	138, 82	≤0.9		0.6
Lanthanum	139, 82	1020:	200:	
Cerium	140, 82	400	190	
Praseodymium	141, 82	1070	630	
Neodymium	142, 82	250	200	
Samarium	144, 82	410	260	
Europium		1910	640	130
Gadolinium		810	340	
Dysprosium		760	460	
Lead	208, 126	1500:	1500:	

Clearly, some kind of element synthesis has been taking place in the peculiar *A* stars (Fo55a, Bu57). Since it is unlikely that the products of any nuclear reaction in the interior could have mixed to the surface, and since an energy source is available in the atmosphere in the form of the magnetic fields, we have suggested that this represents a special process of element synthesis which can only take place in stellar atmospheres where there is acceleration of nucleons by electromagnetic activity to energies far above thermal. The predominance of magic-number nuclei among the overabundant elements indicates that the type of synthesis probably involves neutrons; in fact, it may be essentially an *s* process taking place in a proton atmosphere. However, the buildup may be by deuteron stripping reactions (d,p) as well as or instead of by (n,γ) processes; (α,n), $(\alpha,2n)$, and $(\alpha,3n)$ may also occur. Free neutrons are thought to be produced by (p,n) processes on the light nuclei, and their presence in a hydrogen atmosphere means that equilibrium abundances of both deuterium and He^3 will be produced. Evidence for the presence of deuterium in magnetic stars is at present lacking; evidence for the possible presence of He^3 in one of them, 21 Aquilae, has been discussed by Burbidge and Burbidge (Bu56c).

The activity in magnetic stars is thought to take place in "spot" regions, with a duration of only a few seconds (the time-scale is limited by the rate of energy loss by bremsstrahlung), but must continue over long time-scales comparable with the lifetime of the stars in order to build the observed anomalies of the heavy elements, so that each gram of the surface layer is processed many times. An important condition for the detection of the anomalies is that no mixing between the surface layer and the interior should occur, or it would dilute the products of nuclear activity so that they would not be detectable. This condition happens to be fulfilled in the *A*-type stars, where the convective layer extends from the photosphere inwards to a depth of only 1000 km (cf. Fo55a).

XII. GENERAL ASTROPHYSICS

A. Ejection of Material from Stars and the Enrichment of the Galaxy in Heavy Elements

Table XII,1 gives a tentative assessment of the production of elements in the Galaxy, the table being subdivided in accordance with the various processes described in the previous sections.

The values given in the third column were obtained in the following way. The fractions by weight of the various groups of elements in the solar system have already been given in Table II,1. These fractions were reduced by a factor 2 (except for hydrogen) to make allowance for the lower concentrations of the heavy elements, relative to the sun, in the stars which are believed to comprise the main mass of the Galaxy. It is clear that some such reduction must be made, and the factor 2, although somewhat uncertain, is based on the work of Schwarzschild *et al.* (Sc57b; cf. also earlier work mentioned therein). The resulting fractions, when taken together with a total mass of $7\times10^{10}\ M_\odot$ for the Galaxy (Sc56), yield the values shown in the third column of Table XII,1.

The products of the synthesizing processes must be distributed in space. Emission from stellar atmospheres apart, two ejection mechanisms are mentioned in the fourth column of Table XII,1, supernovae, and red giants and supergiants. These are believed to be of main importance, but matter is also ejected from novae, close binary systems (St46, Wo50, St57a), P Cygni and WR stars, and planetary nebulae (cf. Gu54, St57). Leaving aside the *x* process, the association between the second and fourth columns of Table XII,1 was made on the basis that only in supernovae are temperatures of thousands of millions of degrees reached. Reasons for this view have been given in Sec. III,F. In contrast, temperatures of hundreds of millions of degrees may be expected in red giants and supergiants (Sec. V, B), and in the other contributors, novae, binaries, etc. Thus, processes of synthesis requiring temperatures of order 10^8 degrees have been associated with the giants, whereas processes requiring temperatures of order 10^9 degrees have been associated with supernovae.

Estimates of the total amount of material which may have been distributed in space by supernovae and red giants and supergiants, assuming a constant

TABLE XII,1.

Elements	Mode of production	Total mass in galaxy ($M_\odot$ as unit)	Astrophysical origin	Total mass of all material ejected over lifetime of galaxy ($M_\odot$ as unit)	Required efficiency
He	H burning	8.1×10^9	Emission from red giants and supergiants	2×10^{10}	0.4
D	x process?	7.5×10^6?	Stellar atmospheres? Supernovae?	?	?
Li, Be, B	x process	8.5×10^2	Stellar atmospheres	?	?
C, O, Ne	He burning	4.3×10^8	Red giants and supergiants	2×10^{10}	2×10^{-2}
Silicon group	α process	4.0×10^7	Pre-Supernovae	2×10^8	0.2
Silicon group	s process	8.5×10^6	Red giants and supergiants	2×10^{10}	4×10^{-4}
Iron group	e process	2.4×10^7	Supernovae	2×10^8	0.1
$A>63$	s process	4.5×10^4	Red giants and supergiants	2×10^{10}	2×10^{-6}
$A<75$	r process	5×10^4	Supernovae Type II	1.7×10^8	3×10^{-4}
$A>75$	r process	10^4	Supernovae Type I	3×10^7	3×10^{-4}
$A>63$	p process	1.3×10^2	Supernovae Type II	1.7×10^8	10^{-6}

rate of star formation and death during the lifetime of the Galaxy, are given in the fifth column of Table XII,1 and were made in the following way.

The rates of supernovae were taken to be 1 per 300 years for Type I and 1 per 50 years for Type II, as estimated by Baade and Minkowski (Ba57a), and the age of the Galaxy was taken as 6×10^9 years. The average mass emission per supernova is not easy to estimate (see Sec. XII,C), but we have taken it to be 1.4 $M_\odot$. The exponential-decay light curve is characteristic of Type I and not of Type II supernovae; this has been taken to indicate that the rapid production and capture of neutrons, leading to synthesis of the heaviest r-process elements, occurs only in Type I supernovae. The lighter r-process elements are probably made in Type II supernovae, while the products of the α and e processes may perhaps be made in either type.

The estimate for the ejection of mass by red giants and supergiants was based on the assumption that such stars shed most of their material into space during the giant phases of their evolution. Thus Deutsch (De56) has shown that matter is streaming rapidly outwards from the surface of an M-type supergiant, and he considers from the spectroscopic data that probably all late-type supergiants are also ejecting material at a comparably rapid rate. There is also some evidence that normal giant stars are ejecting matter, although probably on a smaller scale. Hoyle (Ho56b) has indicated that mass loss may have a more important effect on the evolution of giants and supergiants of very low surface gravity than have nuclear processes. With these considerations in mind, it is necessary to estimate the fraction of the mass of the Galaxy that has been condensed into stars which have gone through their whole evolutionary path.

The luminosity and mass functions for solar-neighborhood stars have been studied by Salpeter (Sa55a). He has shown that stars which lie on the main sequence now with absolute visual magnitudes fainter than +3.5 (which form the bulk of stars at present in existence) make up 55% of the total mass. Stars which formed with magnitudes brighter than +3.5 (the majority of which have gone through their whole evolutionary path) make up the remaining 45%. We find from his mass function that the average mass of the stars brighter than +3.5 was 4 $M_\odot$. Since the upper limit to the mass of a white dwarf is about 1.4 $M_\odot$, we find that, of the remaining 45%, 16% may lie in white dwarfs and 29% in the form of ejected gas. Thus the division of mass between stars which have not evolved, white dwarfs, and material which has been processed by stars is approximately 4:1:2. Extrapolating these results to the Galaxy as a whole, we find that 2×10^{10} $M_\odot$ has been processed by stars. Some of this is included in the estimate of matter ejected by supernovae, but the majority will have been processed at temperatures $\lesssim10^8$ degrees, and ejected by red giants and supergiants.

The efficiency of production of each group of elements in the total ejected material that is necessary to explain the abundances in column 3 of Table XII,1 is obtained by dividing column 3 by column 5, and is given in the last column. It must be emphasized here that the estimates given in Table XII,1 are very preliminary. Even on an optimistic view they might well prove incorrect by as much as a factor 3. It is noteworthy, however, that the emission seems adequate to explain the synthesis requirements, except possibly in the case of deuterium, and helium, which is the only element needing an efficiency near to unity for its production. It will be clear from the foregoing that our estimates are not accurate enough to establish whether or not it is necessary to assume some helium in the original matter of the Galaxy.

The efficiency necessary to produce the r-process elements is considerably smaller than that required to give the observed supernova light curves (see Sec. XII,C). It should, in addition, be reiterated that we have no evidence at all concerning the abundances of the r-only and p-process isotopes outside the solar system, as we pointed out in Sec. XI.

If the universal value of the deuterium to hydrogen ratio turns out to be of order 10^{-4}, it will be difficult to provide for an adequate abundance of deuterium by processes of an electromagnetic nature occurring in stellar atmospheres, as has already been pointed out in Sec. X,C. The possible alternative process of origin, also discussed in Sec. X,C, rests on considerations to be given in Sec. XII,B. At sufficiently high temperatures ($T>7\times10^9$ degrees) elements of the iron group break down into α particles and neutrons. For example, $Fe^{56}\rightarrow13\alpha+4n$. In Sec. XII,B the onset of a supernova is associated with such breakdown reactions developing in the central regions of a star. In Type II supernovae, hydrogen is present in the outer expanding envelope, and capture of the neutrons by relatively cool hydrogen yields deuterium without subsequent disintegration. As was pointed out in Sec. X,C, if one supernova converted one solar mass of its envelope into deuterium and was then diluted by mixture with 10^4 solar masses, the terrestrial abundance would result. These estimates of production and dilution are not at all unreasonable. On this basis the galactic abundance of deuterium would not necessarily be as large as in the solar system.

The balance between the observed solar-system abundances and the material ejected from stars has been made on the assumption of a constant rate of star formation and death during 6×10^9 years. There are some considerations which suggest that the rate may have been greater in the early history of the Galaxy.

First, recent work on the 21-cm radiation by hydrogen, reviewed by van de Hulst (Hu56), has shown that neutral atomic hydrogen makes up only a few percent of the total mass of both our Galaxy and M31 (and it probably comprises the major part of the gas and dust). Star formation is currently occurring only in Population I regions, which comprise about 10% of the mass of the Galaxy (by analogy with the ratio found by Schwarzschild (Sc54) in M31). Originally the Galaxy must have been composed entirely of gas. Thus it would seem probable that star formation and element synthesis would have occurred at a greater rate early in the life of the Galaxy, and, as a larger and larger fraction of the mass became tied up in small stars and in white dwarfs, so nuclear activity would have been a decreasing function of the age of the Galaxy.

Second, the solar system is 4.5×10^9 years old, and the uranium isotope ratio (Sec. VIII) gives 6.6×10^9 years as the minimum age of the r-process isotopes. Thus the abundances that we are discussing (which are half the solar system abundances) must have been formed (by all 8 synthesizing processes) in a time less than the age of the Galaxy, say 1–2×10^9 years. Also, there are apparently large abundance differences between extreme Population II stars such as the globular clusters M92 (and M3) and, for example, the old galactic cluster M67, while the ages that have been given do not differ correspondingly and are both $\sim6\times10^9$ years (Jo56, Sa57a). It is of interest to mention recent investigations of the color-magnitude diagram of the nearby stars by Sandage (Sa57b) and by Eggen (Eg57), based on the data of Eggen (Eg55, Eg56) and the Yale Parallax Catalogue (Ya52). There are a few stars to the right of the main sequence at $M=3.5$ to 4; if the parallaxes are accurate then they may be either stars which are still contracting and have not yet reached the main sequence, or they may have evolved off the main sequence. In the latter case, their ages would be $\gtrsim8\times10^9$ years.

These facts may be reconciled if we suppose that star formation and death in the originally spherical gaseous Galaxy occurred at a sufficient rate to produce element abundances which then recondensed into the oldest star systems that we see today in the halo. The initial differentiation of Populations I and II was interpreted as one of location in the Galaxy and of comparative youth (spiral arms) and age (halo, nucleus). However, it is clear that a number of subclassifications are demanded (Pa50, We51, Bu56a). While the oldest halo stars that we see today were condensing, we may suppose that concentration towards the center and the equatorial plane was occurring, and the formation of stars was favored by conditions in the flattening higher-density region, leading to a rapid gradient of heavy-element enrichment so that star systems forming in the disk at nearly the same time as the globular clusters had higher abundances of heavy elements. Only gas will contract in this way; stars, once formed, become "frozen" in the shape of the system at the time of their formation. The remaining gas became concentrated into the spiral arms, so that these were the only places where any star formation could any longer occur. If, during the first 1–2×10^9 years in the life of the Galaxy, the average proportion of the mass in which star formation was occurring is taken as 50% (instead of 10% as now), then the increased rate of nuclear activity would compensate for the shorter time scale for element synthesis.

Finally, it is of some importance to discuss the cosmological background to this problem. In an evolutionary explosive cosmology we have that, if T is the age of a system, $T_{\text{universe}}>T_{\text{galaxy}}>T_{\text{some clusters}}>T_{\text{sun}}$. Apart from the products of an initial ylem, if this type of initial condition is proposed, the Galaxy will condense out of primeval hydrogen, and the synthesis will go on as we have already described it. It is of interest that possibly the ylem production of deuterium could be incorporated in a model of this sort. Alternatively, in a steady-state cosmology we have only that $T_{\text{galaxy}}>T_{\text{some clusters}}>T_{\text{sun}}$, and we should expect that the initial condensation of the Galaxy would take place out of gas which already had a weak admixture of the heavy elements that had been synthesized in other galaxies. Thus it might be that a search for low-mass stars of pure hydrogen would present

a possible cosmological test, although such stars might take so long to condense (He55) that they would have become contaminated with the products of the death of massive stars before they had been able to contract themselves into self-contained stars.

It is also of interest to remark on the final stages of galactic evolution. Although, as remarked in the Introduction, the lowest energy state of the elements would be reached when the galaxy was transmuted to iron, it appears that the astrophysical circumstances lead to the matter being finally all contained in white dwarfs (degenerate matter), so that such an extreme condition as an iron galaxy may never be reached.

It should be emphasized that the present value of $1/H$ determined from the red-shift measurements by Humason, Mayall, and Sandage (Hu56a) is 5.4×10^9 years, and this in the evolutionary cosmologies is a time of the order of the age of the universe. However, this value is in conflict with the ages of some star clusters in our galaxy. Thus any attempt to relate the synthesis of the elements in the galaxy through a number of necessarily complex steps to a particular cosmological model must await the resolution of this dilemma.

B. Supernova Outbursts

In the following discussion we consider the course of evolution that in our view leads to the outbursts of supernovae. According to a well-known calculation by Chandrasekhar the pressure balance in a star cannot be wholly maintained by degeneracy for masses greater than a certain critical mass. For pure He^4 this critical mass turns out to be 1.44 $M_\odot$, while for pure iron the value is 1.24 $M_\odot$.

It follows that stars with masses greater than the critical value cannot be in mechanical support unless there is an appreciably temperature contribution to their internal pressures. Mechanical support therefore demands high internal temperatures in such stars.

Our arguments depend on these considerations. We are concerned with stars above the Chandrasekhar limit, and assume that mechanical support is initially operative to a high degree of approximation. This is not a restriction on the discussion, since our eventual aim will be to show that mechanical support ceases to be operative. A discussion of catastrophic stars would indeed be trivial if a lack of mechanical support were assumed from the outset.

The next step is to realize that a star (with mass greater than the critical value) must go on shrinking indefinitely unless there is some process by which it can eject material into space. The argument for this startling conclusion is very simple. Because of the high internal temperature, energy leaks outwards from the interior to the surface of the star, whence it is radiated into space. This loss of energy can be made good either by a slow shrinkage of the whole mass of the star (the shrinkage being "slow" means that mechanical support is still operative to a high degree of approximation), or by a corresponding gain of energy from nuclear processes. But no nuclear fuel can last indefinitely, so that a balance between nuclear energy generation and the loss from the surface of the star can only be temporary. For stars of small mass the permissible period of balance exceeds the present age of the galaxy, but this is not so for the stars of larger mass now under consideration. Hence for these stars shrinkage must occur, a shrinkage that is interrupted, but only temporarily, whenever some nuclear fuel happens for a time to make good the steady outflow of energy into space.

Shrinkage implies a rising internal temperature, since mechanical support demands an increasing thermal pressure as shrinkage proceeds. It follows that the internal temperature must continue to rise so long as the critical mass is exceeded, and so long as mechanical support is maintained. The nuclear processes consequent on the rise of temperature are, first, production of the α-particle nuclei at temperatures from $1\text{–}3\times10^9$ degrees and, second, production of the nuclei of the iron peak at temperatures in excess of 3×10^9 degrees. It is important in this connection that the temperature is not uniform inside a star. Thus near the center, where the temperature is highest, nuclei of the iron peak may be formed, while outside the immediate central regions would be the α-particle nuclei, formed at a somewhat lower temperature, and still further outwards would be the light nuclei together with the products of the s process, formed at a much lower temperature. Indeed some hydrogen may still be present in the outermost regions of the star near the surface. The operation of the r process turns out to depend on such hydrogen being present in low concentration. The p process depends, on the other hand, on the outer hydrogen being present in high concentration.

Turning now to the nuclei of the iron peak, the statistical equations given in Sec. IV show that the peak is very narrow at $T\sim3\times10^9$ degrees, the calculated abundances falling away sharply as the atomic weight either decreases or increases from 56 by a few units. At higher temperatures the peak becomes somewhat wider, ranging to copper on the upper side and down to vanadium on the lower side. Beyond these limits the abundances still fall rapidly away to negligible values with one exception, the case of He^4. From (6) in Sec. IV we find that the statistical equations yield the following relation between the abundances of He^4 and Fe^{56}

$$\log n(4,2)=34.67+\tfrac{3}{2}\log T_9 +\frac{1}{14}\{\log n(56,26)-36.40-\tfrac{3}{2}\log T_9\}-\frac{34.62}{T_9}+\frac{\theta}{7}$$

or

$$\log n(4,2)-\frac{\log n(56,26)}{14}=32.08+1.39\log T_9-\frac{34.62}{T_9}+\frac{\theta}{7}.$$

Under the conditions of the present discussion the densities of free protons and neutrons are comparable and small compared with the helium and iron densities, so that the θ term is small and is neglected.

We consider the application of this equation for a fixed value of the total density equal to 10^8 g/cm^3. This is a plausible value for $T \simeq 7\times10^9$ degrees ($T_9 \simeq 7$), and is the mean density of a stellar core of $1M_\odot$ with a radius of 1.7×10^8 cm. Supposing that conditions in this core are such that the mass is equally divided between He^4 and Fe^{56}, then log $(4,2)=30.88$ and log $n(56,26)=29.73$, since no other nucleus makes an appreciable contribution to the density. Substituting in the equation above we find that then $T=7.6\times10^9$ degrees. Thus when T rises above about 7×10^9 degrees there is a very considerable conversion of iron into helium.

This result is not very dependent on the particular value of 10^8 g/cm^3 used above, since the critical temperature at which the masses of helium and iron are comparable increases only slowly with the density (Ho46). The density cannot differ very much from 10^8 g/cm^3. Thus a value much less than 10^6 g/cm^3 would be most implausible for $T\sim7\times10^9$ degrees, while a value appreciably greater than 10^8 g/cm^3 is forbidden by the condition that there must be an appreciable thermal contribution to the pressure.

The argument now runs as follows. Let us suppose that the temperature increases to 8.2×10^9 degrees. In this case the right-hand side of the equation increases by 0.4 and in order that the equation is satisfied, the difference between $\log n(4,2)$ and $(1/14)\log n(56,26)$ must also increase by 0.4. However, $\log n(4,2)$ can only increase by 0.3 since at this point the total mass of the core will be in the form of helium. Thus an increase in temperature of 0.6×10^9 degrees implies that the numerical value of the term $(1/14)\log n(56,26)$ is decreased by at least 0.1, and this implies a drop in the amount of iron of at least a factor of 25. So we conclude that an increase in temperature from 7.6 to 8.2×10^9 degrees causes nearly all of the iron to be converted to helium; i.e., from the situation that the mass is equally divided between helium and iron we reach conditions in which $\sim98\%$ of the mass is helium and $\sim2\%$ is iron. A similar conclusion follows for other values of the total density. However, for much higher temperatures the term in θ becomes important, reflecting the approach toward a neutron core which would be achieved at extremes of density and temperature.

To convert 1 g of iron into helium demands an energy supply of 1.65×10^{18} ergs. This may be compared with the total thermal energy of 1 gram of material at 8×10^9 degrees which amounts to only 3×10^{17} ergs. Evidently the conversion of iron into helium demands a supply of energy much greater than the thermal content of the material. The supply must come from gravitation, from a shrinkage of the star, and clearly the shrinkage must be very considerable in order that sufficient energy becomes available. This whole energy supply must go into the conversion and hence into nuclear energy, so that very little energy is available to increase the thermal content of the material. However, instantaneous mechanical stability in such a contraction would demand a very large increase in the internal temperature. Thus we conclude that in this contraction, in which the thermal energy is, by hypothesis, only increased by a few percent, there is no mechanical stability, so that the contraction takes place by free fall inward of the central parts of the star. At a density of 10^8 g/cm^3, this implies an implosion of the central regions in a time of the order of 1/5 of a second ($\tau\simeq(4/3\pi G\rho)^{-\frac{1}{2}}$ and in our view it is just this catastrophic implosion that triggers the outburst of a supernova.

Before considering the consequences of implosion, it is worthwhile emphasizing that loss of energy by neutrino emission, the urca process (Ga41), is very weak compared to the process described above. Neutrino emission does not rob a star of energy at a sufficiently rapid rate to demand catastrophic implosion, though at temperatures in excess of $\sim3\times10^9$ degrees it does promote a much more rapid shrinkage of the star than would otherwise arise from the loss of energy through normal radiation into space. The reaction $Mn^{56}(\beta^-\nu)Fe^{56}$ with a half-life of 2.58 hr probably gives the main contribution to the loss of energy through neutrino emission. The binding energy of Mn^{56} is less than that of Fe^{56} by 2.895 Mev. Thus, under conditions of statistical equilibrium near $T=5\times10^9$ degrees, this implies that Mn^{56} is less abundant than Fe^{56} by a factor $\sim10^3$. Hence, it is reasonable to suppose that a fraction of about 10^{-3} of the mass is in the form of Mn^{56}. The beta decay of this nucleus leads to an energy loss of about 3×10^{12} ergs per gram of Mn^{56}. Thus a time-scale of about 10^8 sec is demanded to produce an energy loss by neutrino emission comparable with the thermal energy possessed by the stellar material. Hence we conclude that under these conditions in the stellar core, the urca process is quite ineffective as compared with the refrigerating action of the conversion of Fe^{56} to He^4.

The last question to be discussed is the relationship between the *implosion* of the central regions of a star and the *explosion* of the outer regions. Two factors contribute to produce explosion in the outer regions. The temperature in the outer regions is very much lower than the central temperature. Because of this the outer material does not experience the same extensive nuclear evolution that the central material does. Particularly, the outer material retains elements that are capable of giving a large energy yield if they become subject to sudden heating, e.g., C^{12}, O^{16}, Ne^{20}, Ne^{21}, He^4, and perhaps even hydrogen. The second point concerns the possibility of the outer material experiencing a sudden heating. Because under normal

conditions the surface temperature of a star is much smaller than the central temperature, material in the outer regions normally possesses a thermal energy per unit mass that is small compared with the gravitational potential energy per unit mass. Hence any abnormal process that causes the thermal energy suddenly to become comparable with the gravitational energy must lead to a sudden heating of the outer material. This is precisely the effect of an implosion of the central regions of a star. Consequent on implosion there is a large-scale conversion of gravitational energy into dynamical and thermal energy in the outer zones of the star.

One last point remains. Will the gravitational energy thus released be sufficient to trigger a thermonuclear explosion in the outer parts of the star? The answer plainly depends on the value of the gravitational potential. Explosion must occur if the gravitational potential is large enough. In a former paper (Bu56) it was estimated that a sudden heating to 10^8 degrees would be sufficient to trigger an explosion. This corresponds to a thermal energy $\sim 10^{16}$ ergs per gram. If the gravitational potential per gram at the surface of an imploding star is appreciably greater than this value, explosion is almost certain to take place. For a star of mass 1.5 $M_\odot$, for instance, the gravitational potential energy per unit mass appreciably exceeds 10^{16} ergs per gram at the surface if the radius of the star is less than 10^{10} cm. At the highly advanced evolutionary state at present under consideration it seems most probable that this condition on the radius of the star is well satisfied. Hence it would appear as if implosion of the central regions of such stars must imply explosion of the outer regions.

Two cases may be distinguished, leading to the occurrence of the *r* and *p* processes. A star with mass only slightly greater than Chandrasekhar's limit can evolve in the manner described above only after almost all nuclear fuels are exhausted. Hence any hydrogen present in the outer material must comprise at most only a small proportion of the total mass. This is the case of hydrogen deficiency that we associate with Type I supernovae, and with the operation of the *r* process. In much more massive stars, however, the central regions may be expected to exhaust all nuclear fuels and proceed to the point of implosion while much hydrogen still remains in the outer regions. Indeed the "central region" is to be defined in this connection as an innermost region containing a mass that exceeds Chandrasekhar's limit. For massive stars the central region need be only a moderate fraction of the total mass, so that it is possible for a considerable proportion of the original hydrogen to survive to the stage where central implosion takes place. This is the case of hydrogen excess that in the former paper we associated with the Type II supernovae, where the *p* process may occur. However, the rarity of the *p*-process isotopes, and hence the small amount of material which must be processed to synthesize them, suggests that if Type II supernovae are responsible, the *p* process is a comparatively rare occurrence even among them. On the other hand, in any supernova in which a large flux of neutrons was produced, a small fraction of those having very high energies might escape to the outer parts of the envelope, and after decaying there to protons, might interact with the envelope material and produce the *p*-process isotopes.

Energy from the explosive thermonuclear reactions, perhaps as much as fifty percent of it, will be carried inwards, causing a heating of the material of the central regions. It is to this heating that we attribute the emission of the elements of the iron peak during the explosion of supernovae.

C. Supernova Light Curves

The left-hand side of Plate 4 shows the Crab Nebula, the remnant of a supernova which exploded in 1054 A.D. The right-hand side of Plate 4 shows three exposures of the galaxy IC 4182 during the outburst and fading of the supernova of 1937. The first exposure (20 min) shows the supernova at maximum brightness; the second (45 min) shows it about 400 days after maximum; on the third (85 min) the supernova is too faint to be detected.

Throughout this paper we have proposed that the production of the *e*-process, *r*-process, and *p*-process isotopes takes place immediately preceding or during such outbursts of supernovae. This is the only type of astrophysical process that we know of, apart from an initial state in an evolutionary cosmology, in which it can be presumed that the necessary high-energy, rapidly evolving conditions can be expected. Thus while we were searching for such an astrophysical situation the discovery that the half-life for decay by spontaneous fission of Cf^{254} was equal to the half-life for decay of the light curve of the supernova in IC 4182 (Bu56, Ba56) led us to make the detailed calculations which are described in Secs. VII and VIII.

Since this original work was carried out, there has been further investigation of the half-life of Cf^{254} (see Sec. VIII). Also we have re-examined the evidence concerning the light curves of supernovae. Results of the calculations have allowed us to re-examine the basic assumption made in our earlier papers, that the energy released in the decay of Cf^{254} must dominate over all of the other possible sources of decay energy.

The half-life of 55±1 days was obtained from Baade's result on the supernova in IC 4128. In Fig. XII, 1 we show Baade's light curve of the supernova in IC 4182, and the light curves of Tycho Brahe's nova and Kepler's nova (Ba43, Ba45), together with estimated points for the supernova of 1054 (the Crab Nebula) (Ma42). For SN IC 4182 the observations have been made for as long as 600 days after maximum. The absolute magnitude of this supernova at maximum has been

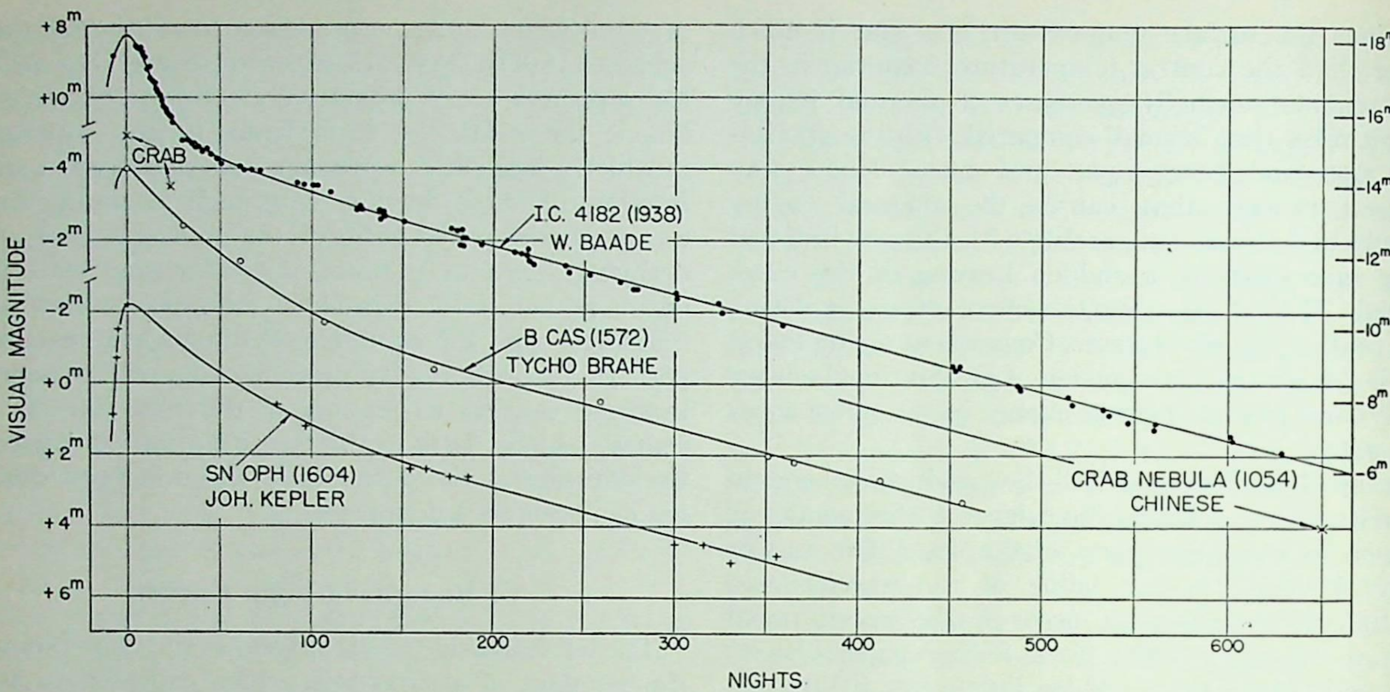


FIG. XII,1. Light curves of supernovae by Baade (Ba43, Ba45, Ba56). Measures for SN IC 4182 are by Baade; those for B Cassiopeiae (1572) and SN Ophiuchi (1604) have been converted by him to the modern magnitude scale from the measures by Tycho Brahe and Kepler. The three points for the supernova of 1054 are uncertain, being taken from the ancient Chinese records (Ma42). The abscissa gives the number of nights after maximum; the left-hand ordinate gives the apparent magnitude (separate scale for each curve); the points for the Crab Nebula belong on the middle scale, i.e., that for B Cassiopeiae. The right-hand ordinate gives the absolute magnitude for SN IC 4182 (Ba57a) derived by using the current distance scale. Compare with Fig. VIII,3.

estimated, with the best current value for the distance scale correction, to be -18.1 (Ba57a). The absolute magnitude of the Crab at its maximum in 1054 has been estimated to be -17.5 to -16.5 (To1355, Ma42, Ba57a). The value of 55 days for the half-life of SN IC 4182 was derived mainly from the points far out in the light curve where apparent magnitudes fainter than $+19$ had to be measured. Any systematic errors in measuring the faint magnitudes in this region would have the effect of changing the estimated value of this half-life and the possibility that such errors are present cannot be ruled out (Ba57a). The uncertainties in the measurements of other supernovae, which are not described here, are such that either all supernovae have true half-lives of 55 days, or they may have a unique half-life slightly different from 55 days, but lying in the range 45–65 days. Alternatively there may be an intrinsic variation between different light curves. These observational uncertainties must be borne in mind when considering the following discussion.

The only comparisons that we shall attempt to make will be between the total radiant energy emitted by a supernova and the form of the light curve, and the total energy emitted and the energy decay curves based on the calculations in Sec. VIII. No direct comparison between the early parts of the supernova light curve and the energy decay curve are possible, since the energy-degradation processes and the energy-transfer processes in the shell of the supernova will distort the relation between the two. Some aspects of this part of the problem have recently been studied by Meyerott and Olds (Me57).

Astrophysical arguments concerning the amounts of the r-process elements which are built in supernovae come from three different directions:

(i) The supernova or the supernovae which synthesized the r-process material of the solar system.
(ii) The Crab Nebula, the only remnant of a supernova which has been studied in any detail.
(iii) The light curves of supernovae which enable us to estimate how much material has been involved in the outburst, either on the assumption that the energy released by one or two of the isotopes dominates, and defines the light curve, or on the assumption that all of the decay activity is important.

Let us suppose that a total mass $mM_\odot$ is ejected in a particular supernova outburst and that a fraction f of this mass is in the form of r-process elements. Further, let us suppose that a fraction g of the heavy elements is in the form of Cf^{254}, or alternatively that a fraction g' of the heavy elements is in the form of Fe^{59}. In Sec. VIII it was shown that under specific conditions this was the only other isotope which could be expected to contribute an amount of energy comparable with Cf^{254}. Since it has a half-life of 45 days it cannot be *a priori* excluded from the discussion, particularly if, as has been stated above, some of the supernova light curves may have half-lives which are near to 45 days.

On the basis that Cf^{254} is responsible for the total

energy released in the light curve after this curve has reached its exponential form, an approximation which will be corrected in what follows, we have, since the energy release per Cf^{254} nucleus is 220 Mev, that the total energy available in the Cf^{254} is $gfm\times1.67\times10^{51}$ ergs. Of the unknowns m, f and g, m is unknown for all supernovae except the Crab which may have a mass of about 0.1 $M_\odot$ (Os57) so that $m\simeq0.1$. The uncertainties in this value are discussed later. The value of f is determined by the conditions in the outer envelope of the star which have developed in its evolution to the supernova stage, while g is determined by the degree of building in the r process.

We consider the possible values of f. The most favorable situation that can be attained is reached when there are equal amounts by number of hydrogen, helium, and carbon, oxygen, and neon taken as a group, and about 1% of iron by number. In schematic terms, the sequence of events is as follows. The protons are captured by the C^{12}, O^{16}, Ne^{20} to form, after beta decay, C^{13}, O^{17}, Ne^{21}. Each of these now captures an alpha particle and releases a single neutron, so that we end with O^{16}, Ne^{20}, and Mg^{24}, together with a source of free neutrons. To build into the heavy element region, about 100 neutrons per Fe^{56} nucleus are required. These neutrons have come, via the steps that we have outlined, from the original hydrogen. Thus by number the necessary ratio of hydrogen to iron is 100:1, and the mass ratios of the material are hydrogen: helium: carbon: oxygen: neon: iron = 1:4:4:5.33:6.67:0.56, so that the iron comprises about 2% of the envelope by mass, and after the outburst about 6% of the mass is transformed into heavy elements. The maximum value of f is then 0.06. This situation is probably unrealistic because we have assumed essentially that the efficiency of the process is 1, since all of the hydrogen has been converted to neutrons, and each light nucleus has produced one neutron. As a more conservative value we will take $f_{max}=0.01$.

A number of values of g are possible; $g=1$ corresponds to the situation in which all of the heavy elements are transmuted to Cf^{254}. We can see no reason from the standpoint of nuclear physics why this should occur and consequently we are inclined to disregard it.

If all of the r-process material is transformed into material lying in the range of atomic weight $230<A<260$, then we see from Fig. VII,4 that $g\simeq0.04$. This is also a very unreal situation, since while no material resides in the region below $A\simeq230$, it is also supposed that no fission has taken place to terminate the r process and hence to begin to cycle the material between $A=110$ and $A=260$.

We can consider a third situation in which, starting from iron, the material has been driven so that all of the r-process peaks have been produced in the way that has been described in Sec. VII. Further nuclear activity results in the following situation. The material begins to cycle between $A=110$ and $A=260$, because, following fission of the heaviest isotopes, the fission fragments help to build in the regions $A=110$ and $A=150$, and the total amount of material which resides in the region $110<A<260$ is determined by the amount of cycling which can take place. Sufficient cycling reduces the amount of material in the region <110, and in the extreme case we suppose that no iron is left and also no material is left in the $N=50$ peak. Under these conditions we estimate, from Figs. VII,3 and VII,4, that $g\simeq0.011$. Here we have supposed that only half of the abundance follows the part of the cycle between $A\simeq110$ and $A\simeq150$.

A fourth situation is just that which has been supposed to give rise to the solar-system abundances of the r-process isotopes. In this case a steady state is reached and no appreciable cycling has taken place. Reference to Fig. VII,3 and VII,4 shows that $g\simeq0.002$.

A fifth situation is that in which the r-process isotopes are built from a base material which is very much lighter than iron. In this case we have only been able to make a guess concerning the amount of material residing in r-process elements with $A\lesssim45$. We then find that $g\simeq0.0004$.

Thus the largest value of g which appears to be plausible is obtained when cycling takes place and $g=0.011$, and it is this value which we use in the following calculations.

There is a spread in absolute magnitudes at maximum of supernovae and probably also a spread in the difference between the maximum absolute magnitude and the magnitude at which the exponential form of the light curve sets in. Let us suppose that this difference is δM. Then the total energy emitted under the exponential tail of the light curve is given by

$$E_t=3.8\times10^{33}\times10^{0.4(4.6-M-\delta M)}\times8.64\times10^4\times\tau_m \text{ ergs},$$

where M is the absolute magnitude of the supernova at maximum and τ_m days is the mean life of the radioactive element responsible. The energy emitted by the sun and its absolute magnitude have been taken as 3.8×10^{33} erg/sec and 4.6, respectively. Thus for Cf^{254},

$$E_t=2\times10^{42}\times10^{0.4(-M-\delta M)}$$

where the half-life of $Cf^{254}=61$ days has been used. If the half-life is 56 days, then the value of E_t must be reduced by about 8%. We have shown in Sec. VIII that the total energy release parallels the spontaneous fission release over the time when the Cf^{254} predominates. Thus when making this calculation we must include the contribution which the other radioactive elements apart from Cf^{254} make to the curve. From Fig VIII,1 we find that the ratio of the total energy release to the Cf^{254} energy release is 8:7. On including this correction factor, we have that the total energy release from the mass $mM_\odot=gfm\times1.91\times10^{51}$ ergs. Thus putting $g=0.011$, $f=0.01$, we have that

$$m\times1.07\times10^{+5}=10^{0.4(-M-\delta M)}$$

or

$$\log m = 0.4(-M-\delta M)-5.029.$$

For the Crab, in which $M \simeq -17.5$ and δM is arbitrarily taken as $+3$,

$$m=5.9,$$

while for SN IC 4182, in which $M=-18.1$ and $\delta M = +2.5$,

$$m=16.1.$$

The mass of the Crab Nebula is estimated to be 0.1 $M_\odot$ (Os57). This result was based on the assumption that, although the hydrogen/helium ratio was abnormally small, in that their number densities were approximately equal, the remainder of the elements had normal abundances, as given, for example, in the appendix. The situation regarding the elements other than hydrogen and helium in the Crab is still very unclear. For example, there appear to be no identifications or information on iron, though we should expect that the iron/hydrogen or iron/helium ratio would be abnormally large. The mass is actually derived by obtaining the density of free electrons per unit of mean atomic weight, and finally by making the assumption about the composition described above. Thus the total mass estimated is approximately a linear function of the mean atomic weight which is assumed. Hence if the material in the shell had the initial composition which we consider to be typical of a highly-evolved star, the mean atomic weight would be about ten times that assumed in Osterbrock's calculation. Hence a mass $\simeq 1\ M_\odot$ would not be unreasonable. This leaves a discrepancy between the calculated and observed mass of this supernova shell of a factor of ~ 5. This may be partly or completely accounted for by uncertainties in estimating the absolute magnitude of the Crab supernova at maximum. Baade (Ba57a) has estimated that the absolute magnitude was -17.5 at maximum by using the observation that the star vanished from naked eye visibility about 650 days after maximum, and assuming that the total drop in brightness was the same as that for the supernova in IC 4182. Mayall and Oort (Ma42) have estimated that it reached -16.5 by using the observation that the supernova ceased to be visible in daylight after 23 days. If the supernova did reach only -16.5 at maximum, the mass needed to be ejected in the envelope is $\sim 2.4\ M_\odot$. An alternative explanation is that the Crab never did show that exponential light curve which is typical of supernovae in which Cf^{254} is produced.

For IC 4182, a mass of $\sim 10\ M_\odot$ is not unreasonable, particularly if it is supposed that the amount of mass ejected is roughly proportional to the maximum luminosity of the supernova.

It is of some interest to carry out a calculation similar to that made for Cf^{254} on the assumption that Fe^{59} is alone responsible for the light curve, particularly since we have earlier pointed out that this isotope will dominate the energy decay among the isotopes apart from Cf^{254}, and also since some supernova light curves have half-lives possibly nearer to 45 than to 55 days.

In this case the total energy released by Fe^{59} is given by

$$E_t = g'fm \times 8.6 \times 10^{49} \text{ ergs}.$$

The most favorable situation which can be envisaged is that as before there is about 2% by mass of iron in the envelope, so that $f \simeq 0.01$, but that there is a paucity of neutrons so that perhaps only 5 neutrons per iron nucleus are made available. On this assumption the proportion of the material which captures 3 neutrons to form Fe^{59} can easily be calculated, as indicated in Sec. VIII, if it is assumed that the capture cross section remains constant, and we find that about 20% of the original Fe^{56} will be transformed to Fe^{59}. In this case $g'=0.2$ (see Sec. VIII). Thus we obtain the result that

$$g'fm \times 4.44 \times 10^{49} = 1.48 \times 10^{42} \times 10^{0.4(M-\delta M)}$$

or

$$\log m = 0.4(-M-\delta M)-4.78.$$

This equation for m has a numerical constant which differs by a factor of 0.25 in the logarithm from that for the Cf^{254} decay. Thus we conclude that if the physical conditions described above are plausible, the large-scale production of Fe^{59} will be able to explain a light curve with a half-life of 45 days if the mass ejected is about 1.8 times greater than that demanded if Cf^{254} is responsible. However, if sufficient neutrons were available to build the r-process elements in the same proportions that they appear in the solar system the relative amount of Fe^{59} can only be estimated rather uncertainly, as has been done in Sec. VIII, but the value of g becomes much smaller than 0.2 and the mass required in the envelope becomes unreasonably large.

The relatively low efficiency of conversion of the material into either Cf^{254} or Fe^{59} arises from the small values of g, g', and f which seem to be plausible. The values of g and g' are determined by the nuclear physics of the situation. However, as previously stated, f is in essence determined by the initial composition of the envelope material. The maximum efficiency would be achieved when $f=1$, i.e., when the whole of the material was initially composed of iron and was changed to heavy elements in the supernova explosion. Such a condition does not seem very plausible. However, if a situation could arise in which some of the outer core material has been converted to isotopes in the iron peak by the equilibrium process (Sec. IV) while the inner core material has evolved to its last equilibrium configuration consisting of a pure neutron core, an instability in which the neutrons were mixed into the iron would result in the production of the r-process isotopes, and since there would be a plentiful supply of neutrons, cycling in the range $110 \lesssim A \lesssim 260$ would

take place. Under these conditions the total amount of mass demanded to produce the curves from Cf^{254} would be only about 0.06 $M_{\odot}$ in the case of the Crab, and only about 0.16 $M_{\odot}$ in the case of SN IC 4182.

Finally, mention must be made of the possible distortion of the light curve by the decay of other isotopes with longer half-lives than those responsible for the regions considered. The most important case appears to be that of Cf^{252}, which has a half-life of 2.2 years. From Fig. VIII,3 this should begin to make the curve flatten out after about 500 days. This flattening is very gradual and could well be masked by the small systematic errors in magnitude measurement which may be present in the light curve of the supernova in IC 4182. If no distortions occur through the energy transfer processes in the envelope at this late stage, very accurate measurements by photoelectric techniques of the very faint end of the next supernova light curve which can be followed may provide a good experimental test of this supernova theory. In the same way further observations may be able to determine whether all supernovae have exactly the same decay curves, a question which may also have some bearing on the problem of whether or not the *r*-process component of the solar-system material was built in a single supernova outburst.

D. Origin of the *r*-Process Isotopes in the Solar System

As mentioned in Sec. VII, the forms of the back sides of the abundance peaks of the *r*-process isotopes might suggest that the conditions obtaining in a single supernova were responsible for their synthesis. Though this conclusion remains highly uncertain at the present time, it is worthwhile considering its astrophysical implications.

A possible sequence of events in the origin of the *r*-process isotopes in the solar system might then be as follows. The isotopes which are made by the other processes have already been synthesized and a cloud of such material is present. Inside this cloud a supernova outburst takes place. About 10^{-2} $M_{\odot}$ is converted by this process to isotopes built in the *r* process, and this material is gradually diluted by the other material present. The material will be diluted first by expansion of the supernova shell into the surrounding medium and second by effects of galactic rotation and turbulence in the interstellar gas. Thus the total volume in which this occurs may be crudely estimated in the following way. The volume of a supernova shell (the Crab Nebula) is about 1 psc^3 after about 10^3 years. Since a shell probably decelerates, it may be supposed that after about 10^4 years its volume is about 25 psc^3 and its expansion velocity has reached a numerical value equal to the mean random velocity of the interstellar clouds.

One time-scale of interest is that for the formation of the solar system, i.e., 10^7-10^8 years, and one epoch of interest is a period of this length occurring some 5×10^9 years ago. However the more important time scale for this argument is the time in which the material can disperse, between the time at which the supernova outburst occurred and the epoch at which the solar system condensed. From the argument based on the U^{235}/U^{238} ratio this time interval is of the order of 10^9 years. At an epoch more than 5×10^9 years ago, it is probable that effects of galactic shear and of turbulent motion were different from their present values, so that no very good estimates of the total dilution volume can be made from this standpoint. On the other hand we can estimate the dilution factor by the following method.

On the basis of discussion in Sec. XII, we can suppose that the *r*-process isotopes originally comprised a total of about 2 percent of a solar mass in the supernova outburst. At present in the solar system the ratio by mass of the *r*-process isotopes to hydrogen is $\sim10^{-6}$ (Table II,1). Thus the amount of hydrogen into which the shell expanded was $\sim2\times10^4$ $M_{\odot}$. If we suppose that the mean density of this gas was about 10^{-24} g/cm^3, the total volume was $\sim4\times10^{61}$ cm^3 so that the dimension of this volume was about 75 psc. This is a fraction of approximately 10^{-6} of the total volume of the galaxy. The total number of supernovae that would explode in $\sim10^9$ years is $\sim3\times10^6$, so that we might expect that, on an average, 3 might have gone off in this volume, so that one only is entirely possible. It does not appear unreasonable from this point of view, therefore, that a single supernova has been responsible for all of the material built by the *r* process currently present in the solar system.

XIII. CONCLUSION

In Secs. III to X we have discussed the details of the various synthesizing processes which are demanded to produce the atomic abundances of all of the isotopes known in nature, while in Secs. II, XI, and XII we have attempted to describe the astrophysical theory and observations which are relevant to this theory.

It is impossible in a short space to summarize the advantages and disadvantages of a theory with as many facets as this. However, it may be reasonable to conclude as follows. The basic reason why a theory of stellar origin appears to offer a promising method of synthesizing the elements is that the changing structure of stars during their evolution offers a succession of conditions under which many different types of nuclear processes can occur. Thus the internal temperature can range from a few million degrees, at which the *pp* chain first operates, to temperatures between 10^9 and 10^{10} degrees when supernova explosions occur. The central density can also range over factors of about a million. Also the time-scales range between billions of years, which are the normal lifetimes of

stars of solar mass or less on the main sequence, and times of the order of days, minutes, and seconds, which are characteristic of the rise to explosion. On the other hand, the theory of primeval synthesis demands that all the varying conditions occur in the first few minutes, and it appears highly improbable that it can reproduce the abundances of those isotopes which are built on a long time-scale in a stellar synthesis theory.

From the standpoint of the nuclear physics of the problem, a major advance in the last few years has been the gradual realization through the interplay between experiment and theory that the helium-burning reaction $3He^4 \rightarrow C^{12}$ will take place in stellar interiors; theoretical work on stellar evolution has shown that in the interiors of red giant stars the conditions are right for such a reaction to take place. Another major advance has been the realization that under suitable stellar conditions the (α,n) reactions on certain nuclei can provide a source of neutrons; these are needed to synthesize the heavy isotopes beyond the iron peak in the abundance distribution and also to build some of the isotopes lighter than this.

The recent analysis of the atomic abundances (Su56) has enabled us to separate the isotopes in a reasonable scheme depending on which mode of synthesis is demanded. In particular, the identification of the *r*-process peaks was followed by the separation of the heavy isotopes beyond iron into the *s*-, *r*-, and *p*-process isotopes, and has enabled us to bring some order into the chaos of details of the abundance curve in this region. The identification of Cf^{254} in the Bikini test and then in the supernova in IC 4182 first suggested that here was the seat of the *r*-process production. Whether this finally turns out to be correct will depend both on further work on the Cf^{254} fission half-life and on further studies of supernova light curves, but that a stellar explosion of some sort is the seat of *r*-process production there seems to be little doubt.

From the observational standpoint, the gradual establishment in the last few years that there are real differences in chemical composition between stars is the strongest argument in favor of a stellar synthesis theory. Details of this argument, and the attempts to show that some stars are going through, or have gone through, particular synthesizing processes, while others are simply condensed out of material which has been processed earlier, and have not yet had time to modify any of their own material, have been given in Sec. XI.

Many problems remain. Although we have given a tentative account of the events leading up to a supernova outburst, we have no detailed dynamical theory of such an explosion. Neither can we explain the principal features of supernova spectra. Further, we do not yet understand the way in which stars evolve to this stage. In fact, the whole problem of stellar evolution beyond the red-giant stage is beset on the theoretical side by problems which are very difficult to handle with the present computational techniques. From the observational side we can estimate the time taken for a star to move through a particular evolutionary stage by making counts of the number of stars in that part of the HR diagram of a star cluster, but these diagrams alone cannot tell us, for example, whether the evolutionary track is in the sense of increasing or decreasing surface temperature at a particular point, i.e., from right to left or from left to right in the diagram.

The whole question of the chemical compositions of stars is complicated by the problem of mixing, since in many cases it cannot be supposed that the atmospheric composition of a star is identical with that of its interior. Also, attempts to determine compositions of very distant or very faint stars are a challenging problem. However, the term "universal abundance" will remain meaningless in the astronomical sense until a reasonable sample of the material in different parts of our own galaxy has been investigated, with the determination of abundances of the heavier elements. It might be pointed out here that on the basis of this theory there may be nothing universal in the isotope ratios. In many cases different isotopes of an element are built by different processes, and the isotope ratios which are observationally known are almost entirely obtained from solar-system material. Thus, for example, another solar system might have condensed out of material consisting mainly of hydrogen and gas ejected from stars which had gone through hydrogen burning, helium burning, the *s* process, the α process, and the *e* process, but not the *r* process. In this case, among the heavy elements the *s*-process isotopes would be present but the *r*-process isotopes would be absent, so that those elements which are predominantly built by the *r* process would be in very low abundance. Thus in such a solar system, the inhabitants would have a very different sense of values, since they would have almost no gold and, for their sins, no uranium.

The question of the composition of the material in other galaxies is not likely to be settled in the near future, except for coarse estimates. Perhaps the only sample of such material which we are able to obtain is contained in the extremely high-energy cosmic-ray particles ($\geq 10^{18}$ ev) which may have been accelerated and escaped from extragalactic nebulae (Ro57, Bu57b).

We have attempted to show that the interchange of material between stars and the interstellar gas and dust will be sufficient to explain the observed abundances relative to hydrogen. This attempt has been reasonably successful, though the aging problems have demanded that we do not consider that synthesis has gone on at a uniform rate since our galaxy condensed. It appears that successive generations of stars are demanded, if all eight processes are to contribute, although it is conceivable that, in a particular star,

one part or another can go through each process, ending with a catastrophic explosion.

We have made some attempt to explain possible modes of production of deuterium lithium, beryllium, and boron, but at present we must conclude that these are little more than qualitative suggestions.

In the laboratory, it is probable that the next steps to be taken are studies of the further reactions in helium burning, i.e., $C^{12}(\alpha,\gamma)O^{16}$ and $O^{16}(\alpha,\gamma)Ne^{20}$, work on the neutron-producing reactions, and particularly on neutron capture cross sections, and accurate work on nuclear masses and binding energies, particularly of the nuclei in the rare-earth region.

In astrophysical observations it is important that evidence for the operation of the *s* process and of the *r* process be sought spectroscopically. In giant stars where the *s* process may operate, the elements most overabundant should be those in general with a stable, "magic number" isotope having a closed shell of neutrons. These elements are strontium, yttrium, zirconium, barium, lathanum, cerium, praseodymium, neodymium, lead, and bismuth. All of these except bismuth have been found to be overabundant in the only such star analyzed up to the present time (Table XI,3). In supernova remnants, such as the Crab nebula, the *r* process may have operated to enhance the abundance of the following heavy elements: selenium, bromine, krypton, tellurium, iodine, xenon, caesium, osmium, iridium, platinum, gold, thorium, and uranium. The effect of the *r* process on the abundance of the iron-group elements and of the lighter elements from carbon to calcium is very difficult to estimate. The iron-group elements will be depleted by neutron capture in the original envelope material but material from the core (*e* process) injected during the explosion may yield an over-all enhancement in the final expanding shell. The over-all abundance of the lighter elements is not appreciably changed by the supernova explosion but the ratio to hydrogen and helium will rise to a value considerably greater than that in unevolved matter. This is because in our model we require that hydrogen and helium originally be comparable in abundance by number to carbon, oxygen, and neon, and then that they be depleted by the energy generation and the neutron-releasing processes.

An important key to the solution of problems in stellar nucleogenesis lies in the determination of relative isotopic abundances in stars. It is realized that this is a very difficult problem spectroscopically, but nonetheless every possible technique of isotope analysis should be brought to bear in this regard. Relative isotopic abundances are many times as informative as relative element abundances.

In concluding we call attention to the minimum age of the uranium isotopes derivable from their relative production in the *r* process. Our calculations place this minimum age at 6.6×10^9 years with a limit of error on the low side of at most 0.6×10^9 years. This small limit of error comes directly from the relatively short half-life of U^{235}. We feel that this minimum age is significantly greater than the currently accepted value for the geochemical age of the solar system, 4.5×10^9 years. It is, indeed, comparable with the ages assigned to the oldest globular clusters so far discovered. Since it is a minimum, it will not be unexpected if still older objects exist in our galaxy. In fact the stars discussed by Sandage and Eggen, mentioned in Sec. XII A, whose ages may be $\gtrsim 8\times10^9$ years, might be such examples.

ACKNOWLEDGMENTS

This work would not have been possible without the help, advice, and criticism of a host of physicists and astronomers. Much unpublished information from many workers has been incorporated into the body of the paper. We are deeply indebted to the following: L. H. Aller, Walter Baade, Rosemary Barrett, W. P. Bidelman, R. F. Christy, G. Dessauer, H. Diamond, Leo Goldberg, J. L. Greenstein, J. A. Harvey, C. Hayashi, R. E. Hester, J. R. Huizenga, R. W. Kavanagh, W. A. S. Lamb, C. C. Lauritsen, T. Lauritsen, N. H. Lazar, W. S. Lyon, R. L. Macklin, P. W. Merrill, R. E. Meyerott, W. Miller, R. Minkowski, F. S. Mozer, Guido Münch, D. E. Osterbrock, R. E. Pixley, A. Pogo, E. E. Salpeter, Allan Sandage, Maarten Schmidt, N. Tanner, Stanley Thompson, and R. Tuttle.

One of the authors (WAF) is grateful for a Fulbright Lectureship and Guggenheim Fellowship at the University of Cambridge in 1954–1955 at which time the studies resulting in this paper were begun.

We also wish to thank Evaline Gibbs and Jan Cooper for typing a difficult manuscript with their usual consummate skill.

APPENDIX

We give in the appendix all of the information that we have been able to collect which is relevant to the synthesis problem. A detailed explanation of the format of this table is given in Sec. II C. The abundances (N) of the isotopes have been taken predominantly from the work of Suess and Urey (Su56), though in a few cases the solar abundances of Goldberg *et al.* (Go57) have been given in parentheses below those of Suess and Urey. The neutron capture cross sections (σ) have been discussed in Sec. V B, while the method of assignment of the nuclei among the eight synthesizing processes has been described in Sec. II B. The designation *m* means "magic," i.e., the nucleus has a closed shell of neutrons and thus a small neutron capture cross section in the *s* process. The designation *m'* means that the progenitor in the *r* process is "magic." When *m* is used as a superscript, as in Eu^{152m}, it designates an isomeric state. A process which returns material back down the neutron capture chain is designated by the word *returns*.

A	Main line	$\sigma(n,\gamma)$ in mb or $\tau_{1/2}$	N	σN	Process	A	Weak line or bypassed	$\sigma(n,\gamma)$ in mb or $\tau_{1/2}$	N	Process
1	${}_{1}H$	0.3(0.1%th)	4.00×10^{10}		Primeval					
2	${}_{1}H_{1}$	$\lesssim 2\times10^{-4}$ (0.1%th)	$5.7\ \times10^{6}$		x					
3	${}_{1}H_{2}(\beta^{-})$	12.26 yr			τ_{β} too long					
3	${}_{2}He_{1}$	5000 (n,p) (extr.)			H burning					
4	${}_{2}He_{2}$		3.08×10^{9}		H burning					
5	${}_{2}He_{3}(n)$	2×10^{-21} sec			returns					
6	${}_{2}He_{4}(\beta^{-})$	0.82 sec								
6	${}_{3}Li_{3}$	$2000(n,\alpha)$(exp)	7.4		x, returns					
7	${}_{3}Li_{4}$	0.33(1%th)	92.6		x					
8	${}_{3}Li_{5}(\beta^{-})$	0.84 sec								
8	${}_{4}Be_{4}(\alpha)$	$\sim2\times10^{-16}$ sec			returns					
9	${}_{4}Be_{5}$	0.085(1%th)	20		x	9	${}_{3}Li_{6}(\beta^{-})\rightarrow n,\ 2He^{4}$	0.17 sec		returns
10	${}_{4}Be_{6}(\beta^{-})$	2.7×10^{6} yr				10	${}_{5}B_{5}$	$1000(n,\alpha)$ (0.1%th)	4.5	returns
11	${}_{4}Be_{7}(\beta^{-})$	Not known								
11	${}_{5}B_{6}$	<0.5(1%th)	19.5		x					
12	${}_{5}B_{7}(\beta^{-})$	0.025 sec								
12	${}_{6}C_{6}$	0.047(1%th)	3.50×10^{6}		He burning					
13	${}_{6}C_{7}$	0.009(1%th)	3.92×10^{4}		H burning					
14	${}_{6}C_{8}(\beta^{-})$	5600 yr								
14	${}_{7}N_{7}$	$1.0(n,\gamma)$(1%th) $1.7(n,p)$(0.1%th)	6.58×10^{6}		H burning					
15	${}_{7}N_{8}$	0.00024(1%th)	2.41×10^{4}		H burning	15	${}_{6}C_{9}(\beta^{-})$	2.4 sec		
16	${}_{7}N_{9}(\beta^{-})$	7.4 sec								
16	${}_{8}O_{8}$	0.01(1%th)	2.13×10^{7}		He burning					
17	${}_{8}O_{9}$	$1.0(n,\alpha)$(0.1%th)	8.00×10^{3}		H burning returns	17	${}_{7}N_{10}(\beta^{-})\rightarrow n,\ O^{16}$	4.14 sec		returns
18	${}_{8}O_{10}$	0.0021(1%th)	4.36×10^{4}		H burning					
19	${}_{8}O_{11}(\beta^{-})$	29 sec								
19	${}_{9}F_{10}$	0.094(1%th)	1600		H burning					
20	${}_{9}F_{11}(\beta^{-})$	11 sec								
20	${}_{10}Ne_{10}$	0.02	7.74×10^{6}		He burning					
21	${}_{10}Ne_{11}$	0.6 (also $n,\ \alpha$)	2.58×10^{4}		H burning returns					
22	${}_{10}Ne_{12}$	0.36(1%th)	8.36×10^{5}	3.0×10^{5}	H burning					
23	${}_{10}Ne_{13}(\beta^{-})$	40 sec								
23	${}_{11}Na_{12}$	3.5	4.38×10^{4}	1.5×10^{5}	H burning, s					
24	${}_{11}Na_{13}(\beta^{-})$	15.0 hr								
24	${}_{12}Mg_{12}$	1.8	7.21×10^{5}	1.3×10^{6}	He burning, α					
25	${}_{12}Mg_{13}$	6 (also $n,\ \alpha$)	9.17×10^{4}	5.5×10^{5}	s					
26	${}_{12}Mg_{14}$	2.0	1.00×10^{5}	2.0×10^{5}	$s(m)$					
27	${}_{12}Mg_{15}(\beta^{-})$	9.5 min								
27	${}_{13}Al_{14}$	5.3 (exp)	9.48×10^{4}	5.0×10^{5}	$s(m)$					
28	${}_{13}Al_{15}(\beta^{-})$	2.30 min								
28	${}_{14}Si_{14}$	$1.8(\frac{1}{3}\times Al^{27})$	9.22×10^{5}	1.7×10^{6}	$\alpha,\ s(m)$					
29	${}_{14}Si_{15}$	12 (also $n,\ \alpha$)	4.70×10^{4}	5.6×10^{5}	s					
30	${}_{14}Si_{16}$	8	3.12×10^{4}	2.5×10^{5}	s					
31	${}_{14}Si_{17}(\beta^{-})$	2.62 hr								
31	${}_{15}P_{16}$	32	1.00×10^{4}	3.2×10^{5}	s					
32	${}_{15}P_{17}(\beta^{-})$	14.5 day								
32	${}_{16}S_{16}$	12 (also $n,\ \alpha$)	3.56×10^{5}	4.3×10^{6}	$\alpha,\ s$					
33	${}_{16}S_{17}$	36 (also $n,\ \alpha$ and $n,\ p$)	2.77×10^{3}	1.0×10^{5}	s					
34	${}_{16}S_{18}$	12	1.57×10^{4}	1.9×10^{5}	s					
35	${}_{16}S_{19}(\beta^{-})$	87 days								
35	${}_{17}Cl_{18}$	33	6670	2.2×10^{5}	s	36	${}_{16}S_{20}$	12	51	$r(m)$
36	${}_{17}Cl_{19}(\beta^{-})$	3.1×10^{5} yr 100				36	${}_{17}Cl_{19}(\beta^{-})$	3.1×10^{5} yr		$\alpha,\ s$
						36	${}_{18}A_{18}$	15	1.26×10^{5} [1.9×10^{6}]	$\alpha,\ s$ s decay
						37	${}_{18}A_{19}$(EC)	35 day		
37	${}_{17}Cl_{20}$	10	2180	2.2×10^{4}	$s(m)$	37	${}_{17}Cl_{20}$	10	2180 [2.2×10^{4}]	$s(m)$
38	${}_{17}Cl_{21}(\beta^{-})$	37.3 min	2.4×10^{4}	4.8×10^{4}	$s(m)$					
38	${}_{18}A_{20}$	?								

A	Main line	$\sigma(n,\gamma)$ in mb or $\tau_{1/2}$	N	σN	Process	A	Weak line or bypassed	$\sigma(n,\gamma)$ in mb or $\tau_{1/2}$	N	Process
39	${}_{18}A_{21}(\beta^-)$	260 yr				39	${}_{18}A_{21}(\beta^-)$	260 yr		
39	${}_{19}K_{20}$	8	2940	2.4×10^4	$s(m)$	40	${}_{18}A_{22}$	15	($+K^{40}$ decay)	s
										s decay
40	${}_{19}K_{21}$	1.3×10^9 yr	0.38		s only	41	${}_{18}A_{23}(\beta^-)$	1.82 hr		
	(EC, 11%,	100								
	β^- 89%)									
41	${}_{19}K_{22}$	25	219	5.5×10^3	s	41	${}_{19}K_{22}$	25	219	s
									[5.5×10^3]	
42	${}_{19}K_{23}(\beta^-)$	12.5 hr				40	${}_{20}Ca_{20}$	4	4.75×10^4	$\alpha(m)$
									[1.9×10^5]	s decay
42	${}_{20}Ca_{22}$	12	314	3.8×10^3	s					
43	${}_{20}Ca_{23}$	20	64	1.3×10^3	s					
44	${}_{20}Ca_{24}$	8	1040	8.3×10^3	α, s					
45	${}_{20}Ca_{25}(\beta^-)$	160 days								
45	${}_{21}Sc_{24}$	16	2.8	45	s					
			(63)	(1.0×10^3)						
46	${}_{21}Sc_{25}(\beta^-)$	85 days								
46	${}_{22}Ti_{24}$	10	194	1.9×10^3	s only	46	${}_{20}Ca_{26}$	4	1.6	r only
			(290)	(2.9×10^3)	e					
47	${}_{22}Ti_{25}$	12	189	2.3×10^3	s					
			(280)	(3.4×10^3)	e, r					
48	${}_{22}Ti_{26}$	6	1790	10.7×10^3	α, s	48	${}_{20}Ca_{28}$	4	87.7	r only (m)
			(2600)	(1.6×10^4)	e, r					
49	${}_{22}Ti_{27}$	8	134	1.1×10^3	s					
			(200)	(1.6×10^3)	e, r					
50	${}_{22}Ti_{28}$	4	130	5.2×10^2	$s(m)$	50	${}_{23}V_{27}$	10	0.55	e
			(190)	(7.6×10^2)	r, e				(1.1)	
51	${}_{22}Ti_{29}(\beta^-)$	5.80 min				50	${}_{24}Cr_{26}$	8	344	e
									(1800)	
51	${}_{23}V_{28}$	30	220		$e(m)$					
			(430)							
52	${}_{23}V_{29}(\beta^-)$	3.77 min								
52	${}_{24}Cr_{28}$	8	6510		$e(m)$					
			(3.3×10^4)							
53	${}_{24}Cr_{29}$	31	744		e					
			(3800)							
54	${}_{24}Cr_{30}$	8	204		e	54	${}_{26}Fe_{28}$	29	3.54×10^4	$e(m)$
			(1000)						(1.3×10^4)	
55	${}_{24}Cr_{31}(\beta^-)$	3.6 min								
55	${}_{25}Mn_{30}$	124	6850		e					
			(7900)							
56	${}_{25}Mn_{31}(\beta^-)$	2.58 hr								
56	${}_{26}Fe_{30}$	29	5.49×10^5		e					
			(2.1×10^5)							
57	${}_{26}Fe_{31}$	96	1.35×10^4		e					
			(5100)							
58	${}_{26}Fe_{32}$	29	1980		e	58	${}_{28}Ni_{30}$	40	1.86×10^4	e
			(760)						(1.7×10^4)	
59	${}_{26}Fe_{33}(\beta^-)$	45 day								
59	${}_{27}Co_{32}$	248	1800		e					
			(2200)							
60	${}_{27}Co_{33}(\beta^-)$	5.2 yr								
60	${}_{28}Ni_{32}$	40	7170		e					
			(6600)							
61	${}_{28}Ni_{33}$	104	342		e					
			(310)							
62	${}_{28}Ni_{34}$	40	1000		e					
			(910)							
63	${}_{28}Ni_{35}(\beta^-)$	80 yr				63	${}_{28}Ni_{35}(\beta^-)$	80 yr		
63	${}_{29}Cu_{34}$	48	146	7.0×10^3	s	64	${}_{28}Ni_{36}$	40	318	s
			(1500)	(7.2×10^4)					[1.3×10^4]	
									(290)	
									[(1.2×10^4)]	
64	${}_{29}Cu_{35}$	12.8 hr				64	${}_{29}Cu_{35}$ (β^- 39%)	12.8 hr		
	(EC 42%;									
	β^+ 19%)									
64	${}_{28}Ni_{36}$	40(0.61)	318	7.8×10^3	s	64	${}_{30}Zn_{34}$	49(0.39)	238	s only
			(290)	(7.1×10^3)					[4.5×10^3]	
65	${}_{28}Ni_{37}(\beta^-)$	2.56 hr				65	${}_{30}Zn_{35}$ (EC, β^+)	245 day		
65	${}_{29}Cu_{36}$	48	66	3.2×10^3	s	65	${}_{29}Cu_{36}$	48	66	s
			(680)	(3.3×10^4)					[3.2×10^3]	
									(680)	
									[(3.3×10^4)]	

A	Main line	σ(n,γ) in mb or $\tau_{1/2}$	N	σN	Process	A	Weak line or bypassed	σ(n,γ) in mb or $\tau_{1/2}$	N	Process
66	${}_{29}Cu_{37}(\beta^-)$	5.1 min								
66	${}_{30}Zn_{36}$	49	134	6.6×10^3	s					
67	${}_{30}Zn_{37}$	64	20.0	1.3×10^3	s					
68	${}_{30}Zn_{38}$	49	90.9	4.5×10^3	s					
69	${}_{30}Zn_{39}(\beta^-)$	52 min								
69	${}_{31}Ga_{38}$	84	6.86	576	s					
70	${}_{31}Ga_{39}(\beta^-)$	21 min								
70	${}_{32}Ge_{38}$	82	10.4	849	s only	70	${}_{30}Zn_{40}$	49	3.35	r only
71	${}_{32}Ge_{39}(EC)$	12 day								
71	${}_{31}Ga_{40}$	84	4.54	381	s					
72	${}_{31}Ga_{41}(\beta^-)$	14.1 hr								
72	${}_{32}Ge_{40}$	82	13.8	1130	s					
73	${}_{32}Ge_{41}$	148	3.84	568	s					
74	${}_{32}Ge_{42}$	82	18.65	1520	s	74	${}_{34}Se_{40}$	96	0.649	p
75	${}_{32}Ge_{43}(\beta^-)$	82 min								
75	${}_{33}As_{42}$	240	$4.0(\frac{3}{5})$	640	sr					
76	${}_{33}As_{43}(\beta^-)$	26.7 hr								
76	${}_{34}Se_{42}$	96	6.16	591	s only	76	${}_{32}Ge_{44}$	82	3.87	r only
77	${}_{34}Se_{43}$	360	5.07		rs					
78	${}_{34}Se_{44}$	96	16.0		r(m′)	78	${}_{36}Kr_{42}$	107	0.175	p
79	${}_{34}Se_{45}(\beta^-)$	7×10^4 yr				79	${}_{34}Se_{45}(\beta^-)$	7×10^4 yr		
		360								
80	${}_{34}Se_{46}$	96	33.8		r(m′)	79	${}_{35}Br_{44}$	480	6.78	r(m′) s decay
81	${}_{34}Se_{47}(\beta^-)$	18 min				80	${}_{35}Br_{45}$ (β^- 92%; EC 5%; β^+ 3%)	18 min		
81	${}_{35}Br_{46}$	480	6.62		r(m′)	80	${}_{36}Kr_{44}$	107	1.14 [122]	s only
82	${}_{35}Br_{47}(\beta^-)$	36 hr				81	${}_{36}Kr_{45}(EC)$	2×10^5 yr 440		
82	${}_{36}Kr_{46}$	107	5.90	630	s only	82	${}_{36}Kr_{46}$	107	5.90 [630]	s only
83	${}_{36}Kr_{47}$	440	5.89		r(m′)	82	${}_{34}Se_{48}$	96	5.98	r only (m′)
84	${}_{36}Kr_{48}$	107	29.3		r(m′)	84	${}_{38}Sr_{46}$	48	0.106	p
85	${}_{36}Kr_{49}(\beta^-)$	10.4 yr								
85	${}_{37}Rb_{45}$	112	4.73		rs(m′)					
86	${}_{37}Rb_{49}(\beta^-)$	18.6 day								
86	${}_{38}Sr_{48}$	48	1.86	89	s only	86	${}_{36}Kr_{50}$	2.4	8.94	r only (m)
87	${}_{38}Sr_{49}$	60	1.33	80	s only	87	${}_{37}Rb_{50}(\beta^-)$	4.3×10^{10} yr 9	1.77	r only (m)
88	${}_{38}Sr_{50}$	4.8	15.6	75	s(m)					
89	${}_{38}Sr_{51}(\beta^-)$	54 day								
89	${}_{39}Y_{50}$	19	8.9	170	s(m)					
90	${}_{39}Y_{51}(\beta^-)$	64.0 hr								
90	${}_{40}Zr_{50}$	12	28.0	336	s(m)					
91	${}_{40}Zr_{51}$	24	6.12	147	s					
92	${}_{40}Zr_{52}$	18	9.32	118	s	92	${}_{42}Mo_{50}$	24	0.364	p(m)
93	${}_{40}Zr_{53}(\beta^-)$	9×10^5 yr				93	${}_{40}Zr_{53}(\beta^-)$	9×10^5 yr		
		24								
94	${}_{40}Zr_{54}$	18	9.48	121	s	93	${}_{41}Nb_{52}$	160	1.00	r, s decay
95	${}_{40}Zr_{55}(\beta^-)$	65 day				94	${}_{41}Nb_{53}^{m}(\beta^-)$	6.6 min		
						94	${}_{41}Nb_{53}(\beta^-)$	2×10^4 yr		
95	${}_{41}Nb_{54}(\beta^-)$	35 day				94	${}_{42}Mo_{52}$	240	0.226	p, s
95	${}_{42}Mo_{53}$	324	0.382	124	s	95	${}_{42}Mo_{53}$	324	0.382 [124]	s
96	${}_{42}Mo_{54}$	240	0.401	96	s only	96	${}_{40}Zr_{56}$	6.0	1.53	r only
						96	${}_{44}Ru_{52}$	432	0.0846	p
97	${}_{42}Mo_{55}$	324	0.232	75	s					
98	${}_{42}Mo_{56}$	240	$0.581(\frac{1}{2})$	70	s ≈ r	98	${}_{44}Ru_{54}$	432	0.0331	p
99	${}_{42}Mo_{57}(\beta^-)$	67 hr								
99	${}_{43}Tc_{56}(\beta^-)$	2.1×10^5 yr 120				99	${}_{43}Tc_{56}(\beta^-)$	2.1×10^5 yr		
100	${}_{43}Tc_{57}(\beta^-)$	16 sec				99	${}_{44}Ru_{55}$	1160	0.191	sr s decay
100	${}_{44}Ru_{56}$	432	0.189	82	s only	100	${}_{44}Ru_{56}$	432	0.189 [82]	s only
101	${}_{44}Ru_{57}$	1160	0.253		rs	100	${}_{42}Mo_{58}$	240	0.234	r only

A	Main line	$\sigma(n,\gamma)$ in mb or $\tau_{1/2}$	N	σN	Process	A	Weak line or bypassed	$\sigma(n,\gamma)$ in mb or $\tau_{1/2}$	N	Process
102	$_{44}$Ru$_{58}$	432	0.467($\frac{1}{2}$)	100	$r \approx s$	102	$_{46}$Pd$_{56}$	504	0.0054	p
103	$_{44}$Ru$_{59}(\beta^-)$	40 day								
103	$_{45}$Rh$_{58}$	760	0.214		r					
104	$_{45}$Rh$_{59}(\beta^-)$	42 sec								
104	$_{46}$Pd$_{58}$	504	0.0628	32	s only	104	$_{44}$Ru$_{60}$	432	0.272	r only
105	$_{46}$Pd$_{59}$	880	0.1536		r					
106	$_{46}$Pd$_{60}$	504	0.1839		r	106	$_{48}$Cd$_{58}$	432	0.0109	p
107	$_{46}$Pd$_{61}(\beta^-)$	7×10^6 yr				107	$_{46}$Pd$_{61}(\beta^-)$	7×10^6 yr		
		880				107	$_{47}$Ag$_{60}$	1000	0.134	r s decay
108	$_{46}$Pd$_{62}$	504	0.180		r	108	$_{47}$Ag$_{61}$ (β^-, 98.5%; EC, 1.5%)	2.3 min		
109	$_{46}$Pd$_{63}(\beta^-)$	13.6 hr				108	$_{48}$Cd$_{60}$	432	0.0079	p, s
						109	$_{48}$Cd$_{61}$(EC)	1.3 yr		
109	$_{47}$Ag$_{62}$	1000	0.126		r	109	$_{47}$Ag$_{62}$	1000	0.126	r
110	$_{47}$Ag$_{63}(\beta^-)$	24 sec								
110	$_{48}$Cd$_{62}$	432	0.111	48	s only	110	$_{46}$Pd$_{64}$	504	0.0911	r only
111	$_{48}$Cd$_{63}$	1160	0.114($\frac{1}{2}$)	66	$r \approx s$					
112	$_{48}$Cd$_{64}$	432	0.212($\frac{1}{2}$)	46	$r \approx s$	112	$_{50}$Sn$_{62}$	84	0.0134	p
113	$_{48}$Cd$_{65}$	1160	0.110($\frac{1}{2}$)	64	$r \approx s$	113	$_{48}$Cd$_{65}{}^m(\beta^-)$	5 yr		
						113	↓ $_{49}$In$_{64}$	1320	0.0046	ps
114	$_{48}$Cd$_{66}$	432	0.256($\frac{1}{2}$)	55	$r \approx s$	114	$_{49}$In$_{65}(\beta^-)$	72 sec		
115	$_{48}$Cd$_{67}(\beta^-)$	54 hr				114	$_{50}$Sn$_{64}$	84	0.0090	ps
115	$_{49}$In$_{66}(\beta^-)$	1320 (6×10^{14} yr)	0.105($\frac{1}{2}$)	70	$s \approx r$					
115	$_{49}$In$_{66}{}^m(\beta^-)$	4.5 hr →				115	$_{50}$Sn$_{65}$	280	0.00465	ps
116	$_{49}$In$_{67}(\beta^-)$	13 sec								
116	$_{50}$Sn$_{66}$	84	0.189	16	s only	116	$_{50}$Sn$_{66}$	84	0.189 [16]	s only
117	$_{50}$Sn$_{67}$	280	0.102	29	s	116	$_{48}$Cd$_{68}$	432	0.068	r only
118	$_{50}$Sn$_{68}$	84	0.316	27	s					
119	$_{50}$Sn$_{69}$	280	0.115	32	s					
120	$_{50}$Sn$_{70}$	84	0.433	36	s	120	$_{52}$Te$_{68}$	156	0.00420	p
121	$_{50}$Sn$_{71}(\beta^-)$	~28 hr								
121	$_{51}$Sb$_{70}$	348	0.141	49	s					
122	$_{51}$Sb$_{71}(\beta^-)$	2.8 day								
122	$_{52}$Te$_{70}$	156	0.115	18	s only	122	$_{50}$Sn$_{72}$	84	0.063	r only
123	$_{52}$Te$_{71}$	520	0.0416	22	s only	123	$_{51}$Sb$_{72}$	348	0.105	r only (m')
124	$_{52}$Te$_{72}$	156	0.221	34	s only	124	$_{50}$Sn$_{74}$	84	0.079	r only (m')
						124	$_{54}$Xe$_{70}$	360	0.00380	p
125	$_{52}$Te$_{73}$	520	0.328		$r(m')$					
126	$_{52}$Te$_{74}$	156	0.874		$r(m')$	126	$_{54}$Xe$_{72}$	360	0.00352	p
127	$_{52}$Te$_{75}(\beta^-)$	9.3 hr								
127	$_{53}$I$_{74}$	880	0.80		$r(m')$					
128	$_{53}$I$_{75}(\beta^-$ 95%)	25.0 min				128	$_{53}$I$_{75}$(EC$+\beta^+$ 5%)	25.0 min		
128	$_{54}$Xe$_{74}$	360	0.0764	28	s only	128	$_{52}$Te$_{76}$	156	1.48	$r(m')$
						129	$_{52}$Te$_{77}(\beta^-)$	72 min		
129	$_{54}$Xe$_{75}$	1400	1.050 ←		$r(m')$	129	$_{53}$I$_{76}(\beta^-)$	1.7×10^7 yr		
						130	$_{53}$I$_{77}(\beta^-)$	12.6 hr		
130	$_{54}$Xe$_{76}$	360	0.162	58	s only	130	$_{54}$Xe$_{76}$	360	0.162 [58]	s only
131	$_{54}$Xe$_{77}$	1400	0.850		r	130	$_{52}$Te$_{78}$	156	1.60	r only (m')
						130	$_{56}$Ba$_{74}$	37.2	0.00370	p
132	$_{54}$Xe$_{78}$	360	1.078		r	132	$_{56}$Ba$_{76}$	37.2	0.00356	p
133	$_{54}$Xe$_{79}(\beta^-)$	5.27 day								
133	$_{55}$Cs$_{78}$	480	0.456		rs					
134	$_{55}$Cs$_{79}(\beta^-)$	2.3 yr								
134	$_{56}$Ba$_{78}$	37.2	0.0886	3.3	s only	134	$_{54}$Xe$_{80}$	360	0.420	r only
135	$_{56}$Ba$_{79}$	124	0.241($\frac{2}{3}$)	20	sr					
136	$_{56}$Ba$_{80}$	37.2	0.286	11	s only	136	$_{54}$Xe$_{82}$	4.9	0.358	r only (m)
						136	$_{58}$Ce$_{78}$	82	0.0044	p
137	$_{56}$Ba$_{81}$	124	0.414($\frac{2}{3}$)	34	sr					
138	$_{56}$Ba$_{82}$	10	2.622	26	$s(m)$	138	$_{57}$La$_{81}$(EC, β^-)	57.6 (2×10^{11} yr)	0.0018	p
						138	$_{58}$Ce$_{80}$	82	0.00566	p
139	$_{56}$Ba$_{83}(\beta^-)$	85 min								
139	$_{57}$La$_{82}$	32	2.00	66	$s(m)$					
140	$_{57}$La$_{83}(\beta^-)$	40.2 hr								
140	$_{58}$Ce$_{82}$	19	2.00	38	$s(m)$					

A	Main line	$\sigma(n,\gamma)$ in mb or $\tau_{\frac{1}{2}}$	N	σN	Process	A	Weak line or bypassed	$\sigma(n,\gamma)$ in mb or $\tau_{\frac{1}{2}}$	N	Process
141	$_{58}\text{Ce}_{83}(\beta^-)$	32 day								
141	$_{59}\text{Pr}_{82}$	60	0.40	24	$s(m)$					
142	$_{59}\text{Pr}_{83}(\beta^-)$	19.1 hr								
142	$_{60}\text{Nd}_{82}$	36	0.39	15	s only (m)	142	$_{58}\text{Ce}_{84}$	82	0.250	r only
143	$_{60}\text{Nd}_{83}$	180	0.175	33	s					
144	$_{60}\text{Nd}_{84}$	120	0.344	42	s	144	$_{62}\text{Sm}_{82}$	72	0.0108	$p(m)$
145	$_{60}\text{Nd}_{85}$	180	0.119	21	s					
146	$_{60}\text{Nd}_{86}$	120	0.248	30	s					
147	$_{60}\text{Nd}_{87}(\beta^-)$	11.6 day								
147	$_{61}\text{Pm}_{86}(\beta^-)$	2.6 yr								
147	$_{62}\text{Sm}_{85}$	4800	0.100		rs					
148	$_{62}\text{Sm}_{86}$	1440	0.0748	108	s only	148	$_{60}\text{Nd}_{88}$	120	0.0824	r only
149	$_{62}\text{Sm}_{87}$	4800	0.0920		rs					
150	$_{62}\text{Sm}_{88}$	1440	0.0492	71	s only	150	$_{60}\text{Nd}_{90}$	120	0.0806	r only
151	$_{62}\text{Sm}_{89}(\beta^-)$	80 yr				151	$_{62}\text{Sm}_{89}(\beta^-)$	80 yr		
		4800								
152	$_{62}\text{Sm}_{90}$	1440	0.176 ← – – – –		rs	151	$_{63}\text{Eu}_{88}$	5600	0.0892	rs
					– – – –	–152	$_{63}\text{Eu}_{89}$(EC, 72%; β^-, 28%)	13 yr or 9.2 hr		
153	$_{62}\text{Sm}_{91}(\beta^-)$	47 hr				152	$_{63}\text{Eu}_{89}{}^{m}$($\beta^-$, 78%; EC, 22%)	13 yr or 9.2 hr		
						152	$_{64}\text{Gd}_{88}$	2400	0.00137	ps
						153	$_{64}\text{Gd}_{89}$(EC)	236 day		
153	$_{63}\text{Eu}_{90}$	5600	0.0976		rs	153	$_{63}\text{Eu}_{90}$	5600	0.0976	rs
154	$_{63}\text{Eu}_{91}(\beta^-)$	16 yr								
154	$_{64}\text{Gd}_{90}$	2400	0.0147	35	s only	154	$_{62}\text{Sm}_{92}$	1440	0.150	r only
155	$_{64}\text{Gd}_{91}$	6800	0.101		r					
156	$_{64}\text{Gd}_{92}$	2400	0.141		r	156	$_{66}\text{Dy}_{90}$	1320	0.00029	p
157	$_{64}\text{Gd}_{93}$	6800	0.107		r					
158	$_{64}\text{Gd}_{94}$	2400	0.169		r	158	$_{66}\text{Dy}_{92}$	1320	0.000502	p
159	$_{64}\text{Gd}_{95}(\beta^-)$	18 hr								
159	$_{65}\text{Tb}_{94}$	6000	0.0956		r					
160	$_{65}\text{Tb}_{95}(\beta^-)$	72 day								
160	$_{66}\text{Dy}_{94}$	1320	0.0127	17	s only	160	$_{64}\text{Gd}_{96}$	2400	0.149	r only
161	$_{66}\text{Dy}_{95}$	4400	0.105		r					
162	$_{66}\text{Dy}_{96}$	1320	0.142		r	162	$_{68}\text{Er}_{94}$	1116	0.000316	p
163	$_{66}\text{Dy}_{97}$	4400	0.139		r					
164	$_{66}\text{Dy}_{98}$	1320	0.157		r	164	$_{68}\text{Er}_{96}$	1116	0.00474	p? Too abundant
165	$_{66}\text{Dy}_{99}(\beta^-)$	2.32 hr								
165	$_{67}\text{Ho}_{98}$	2800	0.118		r					
166	$_{67}\text{Ho}_{99}(\beta^-)$	27.2 hr								
		>30 yr								
166	$_{68}\text{Er}_{98}$	1116	0.104		r					
167	$_{68}\text{Er}_{99}$	1840	0.0770		r					
168	$_{68}\text{Er}_{100}$	1116	0.0850		r	168	$_{70}\text{Yb}_{98}$	1080	0.00030	p
169	$_{68}\text{Er}_{101}(\beta^-)$	9.4 day								
169	$_{69}\text{Tm}_{100}$	1400	0.0318		r					
170	$_{69}\text{Tm}_{101}(\beta^-)$	129 day								
170	$_{70}\text{Yb}_{100}$	1080	0.00666	7.2	s	170	$_{68}\text{Er}_{102}$	1116	0.0470	r only
171	$_{70}\text{Yb}_{101}$	1200	0.0316$(\frac{1}{2})$	19	$s\approx r$					
172	$_{70}\text{Yb}_{102}$	1080	0.0480$(\frac{2}{3})$	35	sr					
173	$_{70}\text{Yb}_{103}$	1200	0.0356$(\frac{2}{3})$	29	sr					
174	$_{70}\text{Yb}_{104}$	1080	0.0678$(\frac{2}{3})$	49	sr	174	$_{72}\text{Hf}_{102}$	520	0.00078	p
175	$_{70}\text{Yb}_{105}(\beta^-)$	4.2 day								
175	$_{71}\text{Lu}_{104}$	1040	0.0488$(\frac{2}{3})$	34	sr	176	$_{70}\text{Yb}_{106}$	1080	0.0278	r only
176	$_{71}\text{Lu}_{105}{}^{m}(\beta^-)$	3.7 hr				176	$_{71}\text{Lu}_{105}(\beta^-)$	Large σ (7.5×10^{10} yr)	0.0013	p
176	$_{72}\text{Hf}_{104}$	520	0.0226	12	s only	177	$_{71}\text{Lu}_{106}(\beta^-)$	6.8 day		
177	$_{72}\text{Hf}_{105}$	1040	0.0806$(\frac{2}{3})$	56	sr	177	$_{72}\text{Hf}_{105}$	1040	0.0806	sr
178	$_{72}\text{Hf}_{106}$	520	0.119	62	s					
179	$_{72}\text{Hf}_{107}$	1040	0.0604	63	s					
180	$_{72}\text{Hf}_{108}$	520	0.155	81	s	180	$_{74}\text{W}_{106}$	480	0.0006	p
181	$_{72}\text{Hf}_{109}(\beta^-)$	46 day								
181	$_{73}\text{Ta}_{108}$	1320	0.065	86	s					
182	$_{73}\text{Ta}_{109}(\beta^-)$	112 day								
182	$_{74}\text{W}_{108}$	480	0.13	62	s					
183	$_{74}\text{W}_{109}$	1640	0.070$(\frac{2}{3})$	77	sr					
184	$_{74}\text{W}_{110}$	480	0.15	72	s	184	$_{76}\text{Os}_{108}$	1176	0.00018	p
185	$_{74}\text{W}_{111}(\beta^-)$	74 day								
185	$_{75}\text{Re}_{110}$	2040	0.0500$(\frac{1}{3})$	51	$s\approx r(m')$					

A	Main line	$\sigma(n,\gamma)$ in mb or $\tau_{\frac{1}{2}}$	N	σN	Process	A	Weak line or bypassed	$\sigma(n,\gamma)$ in mb or $\tau_{\frac{1}{2}}$	N	Process
186	${}_{75}Re_{111}$ (β^- ~95%)	91 hr				186	${}_{75}Re_{111}$ (EC ~5%)	91 hr		
186	${}_{76}Os_{110}$	1176	0.0159	19	*s* only	186	${}_{74}W_{112}$	480	0.14	*r(m′)*
						187	${}_{74}W_{113}(\beta^-)$	24 hr		
187	${}_{76}Os_{111}$	2480	0.0164	41	*s* only	187	${}_{75}Re_{112}(\beta^-)$	2040 (~5×10[10] yr)	0.0850	*r(m′)*
						188	${}_{75}Re_{113}(\beta^-)$	17 hr		
188	${}_{76}Os_{112}$	1176	0.133		*r(m′)*	188	${}_{76}Os_{112}$	1176	0.133	
189	${}_{76}Os_{113}$	2480	0.161		*r(m′)*					
190	${}_{76}Os_{114}$	1176	0.264		*r(m′)*	190	${}_{78}Pt_{112}$	384	0.0001	*p*
191	${}_{76}Os_{115}(\beta^-)$	16 day								
191	${}_{77}Ir_{114}$	3160	0.316		*r(m′)*					
192	${}_{77}Ir_{115}$ (β^- 96.5%)	74 day				192	${}_{77}Ir_{115}$ (EC 3.5%)	74 day		
192	${}_{78}Pt_{114}$	384	0.0127	4.9	*s* only	192	${}_{76}Os_{116}$	1176	0.410	*r(m′)*
193	${}_{78}Pt_{115}$	long				193	${}_{76}Os_{117}(\beta^-)$	31 hr		
193	${}_{78}Pt_{115}^{(m)}$(EC)	3.4 day								
193	${}_{77}Ir_{116}$	3160	0.505		*r(m′)*	193	${}_{77}Ir_{116}$	3160	0.505	*r(m′)*
194	${}_{77}Ir_{117}(\beta^-)$	19 hr								
194	${}_{78}Pt_{116}$	384	0.533		*r(m′)*					
195	${}_{78}Pt_{117}$	1240	0.548		*r(m′)*					
196	${}_{78}Pt_{118}$	384	0.413		*r*	196	${}_{80}Hg_{116}$	240	0.00045	*p*
197	${}_{78}Pt_{119}(\beta^-)$	19 hr								
197	${}_{79}Au_{118}$	1200	0.145		*r*					
198	${}_{79}Au_{119}(\beta^-)$	2.70 day								
198	${}_{80}Hg_{118}$	240	0.0285	6.8	*s* only	198	${}_{78}Pt_{120}$	384	0.117	*r* only
199	${}_{80}Hg_{119}$	680	0.0481		*rs*					
200	${}_{80}Hg_{120}$	240	0.0656(⅔)	10½	*sr*					
201	${}_{80}Hg_{121}$	680	0.0375		*rs*					
202	${}_{80}Hg_{122}$	240	0.0844(⅔)	14	*sr*					
203	${}_{80}Hg_{123}(\beta^-)$	48 day								
203	${}_{81}Tl_{122}$	276	0.0319(⅔)	6	*sr*					
204	${}_{81}Tl_{123}$ (β^- ~98%)	4.1 yr				204	${}_{81}Tl_{123}$ (EC ~2%)	4.1 yr		
204	${}_{82}Pb_{122}$	50	0.0063 (0.26)	0.32 (13)	*s* only	204	${}_{80}Hg_{124}$	240	0.0194	*r*
205	${}_{82}Pb_{123}$(EC)	~5×10[7] yr 100				205	${}_{80}Hg_{125}(\beta^-)$	5.2 min		
						205	${}_{81}Tl_{124}$	276	0.0761(⅔) [14]	*sr* *s* decay
						206	${}_{81}Tl_{125}(\beta^-)$	4.20 min		
206	${}_{82}Pb_{124}$	25	0.122(½) (5.0)	1.5 (62)	*s* cycles *r* decay	206	${}_{82}Pb_{124}$	25	0.122(½) (5.0) [1.5] [(62)]	*s* cycles *r* decay
207	${}_{82}Pb_{125}$	50	0.0995(½) (4.1)	2.5 (100)	*s* cycles *r* decay					
208	${}_{82}Pb_{126}$	10	0.243 (10)	2.4 (100)	*s(m)* cycles (*r* decay)					
209	${}_{82}Pb_{127}(\beta^-)$	3.3 hr								
209	${}_{83}Bi_{126}$	15	0.144(?)	1.2(?)	*s(m)* cycle (*r* decay)					
210	${}_{83}Bi_{127}^{m}$ (β^- 56%) (RaE)	5.0 day				210	${}_{83}Bi_{127}$ (α 44%)	(2.6×10[6] yr) 5		
210	${}_{84}Po_{126}(\alpha)$	138.40 day			(*m*)	211	${}_{83}Bi_{128}(\alpha)$	2.15 min		
206	${}_{82}Pb_{124}$	cycles				207	${}_{81}Tl_{126}(\beta^-)$	4.78 min		
						207	${}_{82}Pb_{125}$	cycles		

BIBLIOGRAPHY

(References are arranged and designated by the first two letters of the first-mentioned author's name, and year. Otherwise duplicate designations are distinguished by postscript letters.)

Ad57 R. L. Adgie and J. S. Hey, Nature 179, 370 (1950).

Aj55 F. Ajzenberg and T. Lauritsen, Revs. Modern Phys. 27, 136 (1955).

Al57 Alder, Stech, and Winther, Phys. Rev. 107, 728 (1957).

Al43 L. H. Aller, Astrophys. J. 97, 135 (1943).

Al51 L. H. Aller and P. C. Keenan, Astrophys. J. 113, 72 (1951).

Al54 L. H. Aller, Mem. soc. roy. sci. Liège 14, 337 (1954).

Al56 L. H. Aller, *Gaseous Nebulae* (John Wiley and Sons, Inc., New York, 1956), p. 217.

Al57a L. H. Aller, Preprint for *Handbuch der Physik* (Springer-Verlag, Berlin, 1957).

Al57b L. H. Aller and J. L. Greenstein (private communication).

Al57c Aller, Elste, and Jugaku, Astrophys. J. Suppl. 3, 1 (1957).

Al57d L. H. Aller, Astrophys. J. **125**, 84 (1957).

Al50 R. A. Alpher and R. C. Herman, Revs. Modern Phys. **22**, 153 (1950).

Al53 R. A. Alpher and R. C. Herman, Ann. Rev. Nuclear Sci. **2**, 1 (1953).

Ar53 Arp, Baum, and Sandage, Astron. J. **58**, 4 (1953).

Aw56 M. Awschalom, Phys. Rev. **101**, 1041 (1956).

Ba43 W. Baade, Astrophys. J. **97**, 119 (1943).

Ba45 W. Baade, Astrophys. J. **102**, 309 (1945).

Ba56 Baade, Burbidge, Hoyle, Burbidge, Christy, and Fowler, Publ. Astron. Soc. Pacific **68**, 296 (1956).

Ba57a W. Baade (private communication).

Ba57 H. W. Babcock, *Proceedings of the Stockholm Symposium on Electromagnetic Phenomena in Cosmical Physics* (to be published).

Ba50 C. L. Bailey and W. R. Stratton, Phys. Rev. **77**, 194 (1950).

Be35 L. Berman, Astrophys. J. **81**, 369 (1935).

Be39 H. A. Bethe, Phys. Rev. **55**, 103, 434 (1939).

Bi50 W. P. Bidelman, Astrophys. J. **111**, 333 (1950).

Bi51 W. P. Bidelman and P. C. Keenan, Astrophys. J. **114**, 473 (1951).

Bi52 W. P. Bidelman, Astrophys. J. **116**, 227 (1952).

Bi53 W. P. Bidelman, Astrophys. J. **117**, 25 (1953).

Bi53a W. P. Bidelman, Astrophys. J. **117**, 377 (1953).

Bi54 W. P. Bidelman, Mem. soc. roy. sci. Liège **14**, 402 (1954).

Bi57 W. P. Bidelman, *Vistas in Astronomy*, edited by A. Beer (Pergamon Press, London, 1957), Vol. II, p. 1428.

Bl52 J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley and Sons, Inc., New York, 1952), p. 390.

Bo39 N. Bohr and J. A. Wheeler, Phys. Rev. **56**, 426 (1939).

Bo53 A. Bohr and B. R. Mottelson, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **27**, No. 16 (1953).

Bo48 H. Bondi and T. Gold, Monthly Notices Roy. Astron. Soc. **108**, 252 (1948).

Bo57 Booth, Ball, and MacGregor, Bull. Am. Phys. Soc. Ser. II, **2**, 268 (1957).

Bo54 Bosman-Crespin, Fowler, and Humblet, Bull. soc. sci. Liège **9**, 327 (1954).

Bo54a R. Bouigue, Ann. astrophys. **17**, 104 (1954).

Br37 British Association Mathematical Tables, Vol. VI.

Br52 British Association Mathematical Tables, Vol. X.

Br49 H. Brown, Revs. Modern Phys. **21**, 625 (1949).

Br57 Bromley, Almqvist, Gove, Litherland, Paul, and Ferguson, Phys. Rev. **105**, 957 (1957).

Bu53 W. Buscombe, Astrophys. J. **118**, 459 (1953).

Bu52 W. Buscombe and P. W. Merrill, Astrophys. J. **116**, 525 (1952).

Bu54 G. R. Burbidge and E. M. Burbidge, Publ. Astron. Soc. Pacific **66**, 308 (1954).

Bu55 G. R. Burbidge and E. M. Burbidge, Astrophys. J. Suppl. **1**, 431 (1955).

Bu55a E. M. Burbidge and G. R. Burbidge, Astrophys. J. **122**, 396 (1955).

Bu56 Burbidge, Hoyle, Burbidge, Christy, and Fowler, Phys. Rev. **103**, 1145 (1956).

Bu56a E. M. Burbidge and G. R. Burbidge, Astrophys. J. **124**, 116 (1956).

Bu56b G. R. Burbidge and E. M. Burbidge, Astrophys. J. **124**, 130 (1956).

Bu56c E. M. Burbidge and G. R. Burbidge, Astrophys. J. **124**, 655 (1956).

Bu57 Burbidge, Burbidge, and Fowler, *Proceedings of the Stockholm Symposium of Electromagnetic Phenomena in Cosmical Physics* (to be published).

Bu57a E. M. Burbidge and G. R. Burbidge, Astrophys. J. **126**, 357 (1957).

Bu57b G. R. Burbidge, Phys. Rev. **107**, 269 (1957).

Ca54 A. G. W. Cameron, Phys. Rev. **93**, 932 (1954).

Ca55 A. G. W. Cameron, Astrophys. J. **121**, 144 (1955).

Ch51 J. W. Chamberlain and L. H. Aller, Astrophys. J. **114**, 52 (1951).

Co53 C. D. Coryell, Ann. Rev. Nuclear Sci. **2**, 308 (1953).

Co56 C. D. Coryell, Laboratory for Nuclear Studies, Annual Report (1956).

Co56a Cowan, Reines, Harrison, Kruse, and McGuire, Science **124**, 103 (1956).

Co57 Cook, Fowler, Lauritsen, and Lauritsen, Phys. Rev. **107**, 508 (1957).

Co57a C. L. Cowan, Jr., and F. Reines, Phys. Rev. **107**, 1609 (1957).

Cr53 J. Crawford, Publ. Astron. Soc. Pacific **65**, 210 (1953).

Cu16 R. H. Curtiss, Publs. Observatory Univ. Mich. **2**, 182 (1916).

Da56 R. Davis, Bull, Am. Phys. Soc. Ser. II, **1**, 219 (1956).

De56 A. J. Deutsch, Astrophys. J. **123**, 210 (1956).

Du51 D. B. Duncan and J. E. Perry, Phys. Rev. **82**, 809 (1951).

Eg55 O. J. Eggen, Astron. J. **60**, 401 (1955).

Eg56 O. J. Eggen, Astron. J. **61**, 360 (1956).

Eg57 O. J. Eggen, Astron. J. **62**, 45 (1957).

Fa57 R. A. Farrell (private communication).

Fe54 Feshbach, Porter, and Weisskopf, Phys. Rev. **96**, 448 (1954).

Fi56 Fields, Studier, Diamond, Mech, Inghram, Pyle, Stevens, Fried, Manning, Ghiorso, Thompson, Higgins, and Seaborg, Phys. Rev. **102**, 180 (1956).

Fo47 Fowler, Hornyak, and Cohen (unpublished, 1947). Quoted in W. E. Siri, *Isotopic Tracers and Nuclear Radiations* (McGraw-Hill Book Company, New York, 1949), p. 11.

Fo54 W. A. Fowler, Mem. soc. roy. sci. Liège **14**, 88 (1954).

Fo55 Fowler, Burbidge, and Burbidge, Astrophys. J. **122**, 271 (1955).

Fo55a Fowler, Burbidge, and Burbidge, Astrophys. J. Suppl. **2**, 167 (1955).

Fo56 W. A. Fowler and J. L. Greenstein, Proc. Natl. Acad. Sci. U. S. **42**, 173 (1956).

Fo56a Fowler, Cook, Lauritsen, Lauritsen, and Mozer, Bull. Am. Phys. Soc. Ser. II, **1**, 191 (1956).

Fo57 P. Fong, preprint (1956); Phys. Rev. (to be published).

Fr55 J. H. Fregeau and R. Hofstadter, Phys. Rev. **99**, 1503 (1955).

Fr56 J. H. Fregeau, Phys. Rev. **104**, 225 (1956).

Fu39 Y. Fujita, Japanese J. Astron. Geophys. **17**, 17 (1939).

Fu40 Y. Fujita, Japanese J. Astron. Geophys. **18**, 45 (1940).

Fu41 Y. Fujita, Japanese J. Astron. Geophys. **18**, 177 (1941).

Fu56 Y. Fujita, Astrophys. J. **124**, 155 (1956).

Ga41 G. Gamow and M. Schoenberg, Phys. Rev. **59**, 539 (1941).

Ga43 G. Gamow Astrophys. J. **98**, 500 (1943).

Gh55 A. Ghiorso, UCRL Report No. 2912, and Geneva conference report (1955).

Go37 V. M. Goldschmidt, Skrifter Norske Videnskaps-Acad. Oslo. I. Mat.-Naturv. Kl. No. 4 (1937).

Go57 Goldberg, Aller, and Müller, preprint (1957).

Go54 Good, Kunz, and Moak, Phys. Rev. **94**, 87 (1954).

Go54a K. Gottstein, Phil. Mag. **45**, 347 (1954).

Gr54b A. E. S. Green, Phys. Rev. **95**, 1006 (1954).

Gr40 J. L. Greenstein, Astrophys. J. **91**, 438 (1940).

Gr47 J. L. Greenstein and W. S. Adams, Astrophys. J. **106**, 339 (1947).

Gr51 J. L. Greenstein and R. S. Richardson, Astrophys. J. **113**, 536 (1951).

Gr54 J. L. Greenstein, *Modern Physics for Engineers*, editor L. Ridenour (McGraw-Hill Book Company, Inc., New York, 1954), p. 267.

Gr54a J. L. Greenstein, Mem. soc. roy. sci. Liège **14**, 307 (1954).

Gr54c J. L. Greenstein and E. Tandberg-Hanssen, Astrophys. J. **119**, 113 (1954).

Gr56 J. L. Greenstein, Publ. Astron. Soc. Pacific **68**, 501 (1956).

Gr56a J. L. Greenstein, *Third Berkeley Symposium on Statistics* (University of California Press, Berkeley, 1956), p. 11.

Gr57 J. L. Greenstein, (private communication).

Gr57a J. L. Greenstein and P. C. Keenan (to be published).

Gr55 G. M. Griffiths and J. B. Warner, Proc. Phys. Soc. (London) **A68**, 781 (1955).

Gu54 G. A. Gurzadian, *Dynamical Problems of Planetary Nebulae* (Erivan, USSR, 1954).

Ha50 R. N. Hall and W. A. Fowler, Phys. Rev. **77**, 197 (1950).

Ha56 C. Hayashi and M. Nishida, Progr. Theoret. Phys. **16**, 613 (1956).

Ha56a Halbert, Handley, and Zucker, Phys. Rev. **104**, 115 (1956).

Ha56b Hayakawa, Hayashi, Imoto, and Kikuchi, Progr. Theoret. Phys. **16**, 507 (1956).

Ha56c C. B. Haselgrove and F. Hoyle, Monthly Notices Roy. Astron. Soc. **116**, 527 (1956).

Ha57 Hagedorn, Mozer, Webb, Fowler, and Lauritsen, Phys. Rev. **105**, 219 (1957).

Ha57a E. C. Halbert and J. B. French, Phys. Rev. **105**, 1563 (1957).

Ha57b J. A. Harvey (private communication).

He48 A. Hemmendinger, Phys. Rev. **73**, 806 (1948).

He49 A. Hemmendinger, Phys. Rev. **75**, 1267 (1949).

He49a G. H. Herbig, Astrophys. J. **110**, 143 (1949).

He55 Henyey, Lelevier, and Levée, Publ. Astron. Soc. Pacific **67**, 154 (1955).

He57 L. Heller, Astrophys. J. **126**, 341 (1957).

He57a G. H. Herbig and K. Hunger (unpublished).

Ho46 F. Hoyle, Monthly Notices Roy. Astron. Soc. **106**, 343 (1946).

Ho49 F. Hoyle, Monthly Notices Roy. Astron. Soc. **109**, 365 (1949).

Ho54 F. Hoyle, Astrophys. J. Suppl. **1**, 121 (1954).

Ho55 F. Hoyle and M. Schwarzschild, Astrophys. J. Suppl. **2**, 1 (1955).

Ho56 Hoyle, Fowler, Burbidge, and Burbidge, Science **124**, 611 (1956).

Ho56b F. Hoyle, Astrophys. J. **124**, 482 (1956).

Ho56c R. Hofstadter, Revs. Modern Phys. **28**, 214 (1956).

Hu51 H. Hurwitz and H. A. Bethe, Phys. Rev. **81**, 898 (1951).

Hu55 J. R. Huizenga, Physica **21**, 410 (1955).

Hu55a "Neutron cross sections," by D. J. Hughes and J. A. Harvey, Brookhaven National Laboratory (1955).

Hu56 H. C. van de Hulst, Verslag Gewone Vergader. Afdel. Natuvrk. Koninkl. Ned. Akad. Wetenschap. **65**, No. 10, 157 (1956).

Hu56a Humason, Mayall, and Sandage, Astron. J. **61**, 97 (1956).

Hu56b J. R. Huizenga and J. Wing, Phys. Rev. **102**, 926 (1956).

Hu57 J. R. Huizenga and J. Diamond, Phys. Rev. **107**, 1087 (1957).

Hu57a J. A. Humblet (unpublished).

Hu57b J. R. Huizenga and J. Wing, Phys. Rev. **106**, 91 (1957).

Hu57c K. Hunger and G. E. Kron, Publ. Astron. Soc. Pacific **69**, 347 (1957).

Jo54 H. L. Johnson, Astrophys. J. **120**, 325 (1954).

Jo56 H. L. Johnson and A. R. Sandage, Astrophys. J. **124** 379 (1956).

Jo57 W. H. Johnson, Jr., and V. B. Bhanot, Phys. Rev. **107**, 1669 (1957).

Ka54 Kaplon, Noon, and Racette, Phys. Rev. **96**, 1408 (1954).

Ka56 R. W. Kavanagh, thesis, California Institute of Technology, and private communication.

Ke41 P. C. Keenan and W. W. Morgan, Astrophys. J. **94**, 501 (1941).

Ke42 P. C. Keenan, Astrophys. J. **96**, 101 (1942).

Ke53 P. C. Keenan and G. Keller, Astrophys. J. **117**, 241 (1953).

Ke56 P. C. Keenan and R. G. Teske, Astrophys. J. **124**, 499 (1956).

Ki56 T. D. Kinman, Monthly Notices Roy. Astron. Soc. **116**, 77 (1956).

Kl47 O. Klein, Arkiv Mat. Astron. Fysik **34A**, No. 19 (1947); F. Beskow and L. Treffenberg, Arkiv Mat. Astron. Fysik **34A**, Nos. 13 and 17 (1947).

Ko56 O. Kofoed-Hansen and A. Winther, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **30**, No. 20 (1956).

La57 W. A. S. Lamb and R. E. Hester, Phys. Rev. **107**, 550 (1957).

La57a W. A. S. Lamb and R. E. Hester, Report No. UCRL/4903 (May, 1957).

La57b Lagar, Lyon, and Macklin, Bull. Am. Phys. Soc. Ser. II, **2**, 15 (1957), and private communication.

Le56 T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956).

Ma42 N. U. Mayall and J. H. Oort, Publ. Astron. Soc. Pacific **54**, 95 (1942).

Ma49 M. G. Mayer and E. Teller, Phys. Rev. **76**, 1226 (1949).

Ma57 J. B. Marion and W. A. Fowler, Astrophys. J. **125**, 221 (1957).

Mc40 A. McKellar, Publ. Astron. Soc. Pacific **52**, 407 (1940).

Mc41 A. McKellar, Observatory **64**, 4 (1941).

Mc44 A. McKellar and W. H. Stillwell, J. Roy. Astron. Soc. Canada **38**, 237 (1944).

Mc48 A. McKellar Publs. Dominion Astrophys. Observatory, Victoria, B. C. **7**, 395 (1948).

Me47 P. W. Merrill, Astrophys. J. **105**, 360 (1947).

Me52 P. W. Merrill, Science **115**, 484 (1952).

Me56 P. W. Merrill and J. L. Greenstein, Astrophys. J. Suppl. **2**, 225 (1956).

Me56a P. W. Merrill, Publ. Astron. Soc. Pacific **68**, 70 (1956).

Me57 R. E. Meyerott and J. Olds, paper in preparation.

Mi50 R. Minkowski, Publs. Observatory Univ. Mich. **10**, 25 (1950).

Mu57 C. O. Muehlhause and S. Oleska, Phys. Rev. **105**, 1332 (1957).

Mu57a G. Münch, Astrophys. J., in preparation.

Na56 Nakagawa, Ohmura, Takebe, and Obi, Progr. Theoret. Phys. **16**, 389 (1956).

Ni55 S. G. Nilsson, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **29**, No. 16 (1955).

No53 R. J. Northcott, J. Roy. Astron. Soc. Canada **47**, 65 (1953).

No55 J. H. Noon and M. F. Kaplon, Phys. Rev. **97**, 769 (1955).

No57 Noon, Herz, and O'Brien, Nature **179**, 91 (1957).

Op39 J. R. Oppenheimer and J. S. Schwinger, Phys. Rev. **56**, 1066 (1939).

Op51 E. J. Öpik, Proc. Roy. Irish Acad. **A54**, 49 (1951).

Op54 E. J. Öpik, Mem. soc. roy sci. Liège **14**, 131 (1954).

Os57 D. E. Osterbrock, Publ. Astron. Soc. Pacific **69**, 227 (1957).

Pa50 P. P. Parenago, Astron. Zhur. **27**, 150 (1950).

Pa53 Patterson, Brown, Tilton, and Inghram, Phys. Rev. **92**, 1234 (1953).

Pa55 C. C. Patterson, Geochim. Cosmochim. Acta **7**, 151 (1955).

Pa55a Patterson, Brown, Tilton, and Inghram, Science **121**, 69 (1955).

Pi57 R. E. Pixley (private communication).

Pi57a Pixley, Hester, and Lamb (private communication).

Po47 D. M. Popper, Publ. Astron. Soc. Pacific **59**, 320 (1947).

Re53 F. Reines and C. L. Cowan, Phys. Rev. **92**, 830 (1953).

Re56 Reynolds, Scott, and Zucker, Phys. Rev. **102**, 237 (1956).

Ro50 N. G. Roman, Astrophys. J. **112**, 554 (1950).

Ro52 N. G. Roman, Astrophys. J. **116**, 122 (1952).

Ro57 B. Rossi (private communication).

Ru29 H. N. Russell, Astrophys. J. **70**, 11 (1929).

Sa40 R. F. Sanford, Publ. Astron. Soc. Pacific **52**, 203 (1940).

Sa44 R. F. Sanford, Astrophys. J. **99**, 145 (1944).

Sa52 E. E. Salpeter, Astrophys. J. **115**, 326 (1952).

Sa52a E. E. Salpeter, Phys. Rev. **88**, 547 (1952).

Sa53 E. E. Salpeter, Ann. Rev. Nuclear Sci. **2**, 41 (1953).

Sa54 E. E. Salpeter, Australian J. Phys. **7**, 373 (1954).

Sa55 E. E. Salpeter, Phys. Rev. **97**, 1237 (1955).

Sa55a E. E. Salpeter, Astrophys. J. **121**, 161 (1955).

Sa57 E. E. Salpeter, Phys. Rev. **107**, 516 (1957).

Sa57a A. R. Sandage, Astrophys. J. **125**, 435 (1957).

Sa57b A. R. Sandage (unpublished).

Sc56 M. Schmidt, Bull. Astron. Soc. Netherlands **13**, 15 (1956).

Sc57a R. P. Schuman, unpublished.

Sc54 M. Schwarzschild, Astron. J. **59**, 273 (1954).

Sc57 Schwarzschild, Howard, and Härm, Astrophys. J. **125**, 283 (1957).

Sc57b Schwarzschild, Schwarzschild, Searle, and Meltzer, Astrophys. J. **125**, 123 (1957).

Sn55 T. Snyder *et al.*, Geneva Conference Reports **5**, 162 (1955).

Sm48 J. S. Smart, Phys. Rev. **74**, 1882 (1948).

Sm49 J. S. Smart, Phys. Rev. **75**, 1379 (1949).

Sp49 L. Spitzer, Astrophys. J. **109**, 548 (1949).

Sp55 L. Spitzer and G. B. Field, Astrophys. J. **121**, 300 (1955).

St56 E. J. Stanley and R. Price, Nature **177**, 1221 (1956).

St46 O. Struve, Observatory **66**, 208 (1946).

St57 O. Struve, Sky and Telescope **16**, 262 (1957).

St57a O. Struve and S. S. Huang, Occ. Notices Roy. Astron. Soc. **3**, 161, No. 19 (1957).

Su56 H. E. Suess and H. C. Urey, Revs. Modern Phys. **28**, 53 (1956).

Sw56 C. P. Swann and E. R. Metzger, Bull. Am. Phys. Soc. **1**, 211 (1956); see also Amsterdam Conference Proceedings (1956).

Sw52 P. Swings and J. W. Swensson, Ann. astrophys. **15**, 290 (1952).

Ta57 N. Tanner, Phys. Rev. **107**, 1203 (1957).

Ta57a N. Tanner and R. E. Pixley (private communication).

Te56 R. G. Teske, Publ. Astron. Soc. Pacific **68**, 520 (1956).

Th52 R. G. Thomas, Phys. Rev. **88**, 1109 (1952).

Th54 A. D. Thackeray, Monthly Notices Roy. Astron. Soc. **114**, 93 (1954).

Th57 S. G. Thompson and A. Ghiorso (private communication).

To1355 T'o-t'o Sung-Shih, *History of the Sung Dynasty*, Treatise on Astronomy, paragraph on guest-stars (Po-na edition), Chap. 56, p. 25a.

To49 Tollestrup, Fowler, and Lauritsen, Phys. Rev. **76**, 428 (1949).

Tr55 G. Traving, Z. Astrophys. **36**, 1 (1955).

Tr57 G. Traving, Z. Astrophys. **41**, 215 (1957).

Tu56 Turkevitch, Hamaguchi, and Read, Gordon Conference Report (1956).

Ur56 H. C. Urey, Proc. Natl. Acad. Sci. U. S. **42**, 889 (1956).

Wa55 A. H. Wapstra, Physica **21**, 367, 387 (1955).

Wa55a Way, King, McGinnis, and van Lieshout, Nuclear Level Schemes, National Academy of Sciences, Washington (1955).

Wa55b G. J. Wasserburg and R. J. Hayden, Nature **176**, 130 (1955).

We35 C. F. von Weizsäcker, Z. Physik **96**, 431 (1935).

We38 C. F. von Weizsäcker, Physik. Z. **39**, 633 (1938).

We51 C. F. von Weizsäcker, Astrophys. J. **114**, 165 (1951).

We57 J. Weneser, Phys. Rev. **105**, 1335 (1957).

We57a V. F. Weisskopf, Revs. Modern Phys. **29**, 174 (1957).

Wh41 J. A. Wheeler, Phys. Rev. **59**, 27 (1941).

Wo50 F. B. Wood, Astrophys. J. **112**, 196 (1950).

Wo49 Woodbury, Hall, and Fowler, Phys. Rev. **75**, 1462 (1949).

Wo52 E. J. Woodbury and W. A. Fowler, Phys. Rev. **85**, 51 (1952).

Wu41 K. Würm, Naturwissenschaften **29**, 686 (1941).

Wu57 Wu, Ambler, Hudson, Hoppes, and Hayward, Phys. Rev. **105**, 1413 (1957).

Ya52 *Yale Parallax Catalogue* (Yale University Observatory, 1952), third edition.

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 4 OCTOBER, 1957

Transfer Problems and the Reciprocity Principle

K. M. CASE*

General Atomic, Division of General Dynamics Corporation, San Diego, California, and The Institute for Advanced Study, Princeton, New Jersey

I. INTRODUCTION

In both the theories of radiative equilibrium and neutron diffusion we encounter transfer equations of the form

$$\mathbf{\Omega}\cdot\nabla\psi(\mathbf{r},\mathbf{\Omega})+\sigma_i(\mathbf{r})\psi_i(\mathbf{r},\mathbf{\Omega})=\sum_j\int F_{ij}(\mathbf{\Omega},\mathbf{\Omega}',\mathbf{r})\sigma_j(\mathbf{r})\psi_j(\mathbf{r},\mathbf{\Omega}')d\Omega'+q_i(\mathbf{r},\mathbf{\Omega}). \quad (1)$$

Since the most direct application of the specific results to be obtained will be to neutron diffusion we will use the terminology of that subject. Here $\psi_i(\mathbf{r},\mathbf{\Omega})$ is the angular flux of neutrons of energy i at $\mathbf{r}$ moving in direction $\mathbf{\Omega}$. $\sigma_i(\mathbf{r})$ is the total cross section for neutrons of energy i and F_{ij} is the probability that if a neutron with energy j, direction $\mathbf{\Omega}'$ suffers a collision at $\mathbf{r}$ a neutron of energy i, direction $\mathbf{\Omega}$, will appear at $\mathbf{r}$. $q_i(\mathbf{r},\mathbf{\Omega})$ describes any external neutron source.

While (1) is, without simplifying assumptions, exceedingly complex, a few general statements can be made. These concern the reciprocity principle and questions of uniqueness. The first of these is most important. Besides enabling us to compare different experimental situations and simplifying much of the mathematics it shows, as will be seen, how apparently difficult problems can be solved by relating them to simpler ones.

Unfortunately, even the most elegant proofs[1] have been rather complex. Indeed, there are a number of theorems all of which go by this same name. In the following it will be shown that all these theorems are consequences of a simple identity. The method of proof has a number of advantages. First, the origin of the theorems and their generalizations in more complex situations are directly apparent. Second, the equations used so powerfully by Chandrasekhar[1] to determine the angular densities are obtained at exactly the same step. In a certain sense Chandrasekhar's argument is turned around. Instead of obtaining equations for the reflection and transmission functions and then deriving a reciprocity principle we start from a slightly generalized principle. From this the equations and the specific reciprocity result follow. This permits one to see exactly what the simplifications resulting from the "Principle of Invariance" [1] are and makes possible a simple proof of this latter principle. A last advantage is that we can in this way obtain the fundamental equations for reflection and transmission functions even when reciprocity in a strict sense does not hold.

Uniqueness theorems are particularly useful in connection with equations like Eq. (1). They not only show what must be specified to achieve a unique solution but also suggest certain notations which materially simplify the obtaining of the solution. Such a theorem is proved in Sec. II for a simple situation. The result is essential for the later arguments. Fortunately the proof is quite similar to the proof of the reciprocity identity in its general form which follows in Sec. III. In that section we obtain as special cases various results which have been called "Reciprocity Principles."

In Secs. IV and V we turn to plane problems and derive Chandrasekhar's[1] equations for the reflection function for a half-space. In Sec. VI these equations are used to obtain the Green's function for one velocity neutron diffusion in the same geometry. A specific reciprocity result of Sec. III is used in Sec. VII to obtain the Green's function for two adjacent half-spaces.

In Sec. VIII Chandrasekhar's equations for a slab are obtained as another specialization of the general reciprocity identity together with certain invariance considerations.

II. A UNIQUENESS THEOREM

For simplicity in deriving the general results and in order to be able to carry through the detailed calculations for specific solutions we restrict ourselves in the body of this article to the following equation which is appropriate for one velocity neutron diffusion:

$$\mathbf{\Omega}\cdot\nabla\psi(\mathbf{r},\mathbf{\Omega})+\sigma(\mathbf{r})\psi(\mathbf{r},\mathbf{\Omega})=c(\mathbf{r})\sigma(\mathbf{r})\int f(\mathbf{\Omega},\mathbf{\Omega}')\psi(\mathbf{r},\mathbf{\Omega}')d\Omega'+q(\mathbf{r},\mathbf{\Omega}). \quad (2)$$

Here $c(\mathbf{r})$ is the number of neutrons emitted at $\mathbf{r}$ per neutron collision at $\mathbf{r}$. It is chosen so that

$$\int f(\mathbf{\Omega},\mathbf{\Omega}')d\Omega=1. \quad (3a)$$

We also assume that

$$\int f(\mathbf{\Omega},\mathbf{\Omega}')d\Omega'=1, \quad (3b)$$

* Permanent address: Physics Department, University of Michigan, Ann Arbor, Michigan.

[1] For example, see S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, New York, 1950).

and

$$f(\mathbf{\Omega},\mathbf{\Omega}') \geq 0 \quad \text{all } \mathbf{\Omega}, \mathbf{\Omega}'. \tag{4}$$

(These conditions are satisfied for all interesting physical situations.)

Theorem.—If $c(\mathbf{r})<1$ throughout a region V, the angular density is uniquely determined by the incident angular distribution on the bounding surface S and the sources within V.

Proof.—Let ψ_1, ψ_2 be two solutions of (2) with the same incident angular distribution. The difference $\psi=\psi_1-\psi_2$ satisfies the equations,

$$\mathbf{\Omega}\cdot\nabla\psi+\sigma(\mathbf{r})\psi=c(\mathbf{r})\sigma(\mathbf{r})\int f(\mathbf{\Omega},\mathbf{\Omega}')\psi(\mathbf{r},\mathbf{\Omega}')d\Omega' \tag{5}$$

and

$$\psi(\mathbf{r}_s,\Omega)=0, \quad \text{for } \mathbf{n}_0\cdot\mathbf{\Omega}<0. \tag{6}$$

Here $\mathbf{r}_s$ is an arbitrary point on S where the outward normal is $\mathbf{n}_0$.

If (5) is multiplied by $\psi(\mathbf{r},\mathbf{\Omega})$ and then integrated over V and $\mathbf{\Omega}$ we find on using Gauss' theorem that

$$I_1=I_2, \tag{7}$$

where

$$I_1=\int \mathbf{n}_0\cdot\mathbf{\Omega}\frac{[\psi(\mathbf{r},\mathbf{\Omega})]^2}{2}d\Omega dS, \tag{8}$$

and

$$I_2=\int_V d\mathbf{r}\sigma(\mathbf{r})\int\int d\Omega d\Omega'\psi(\mathbf{r},\mathbf{\Omega})f(\mathbf{\Omega},\mathbf{\Omega}')\psi(\mathbf{r},\mathbf{\Omega}')$$
$$-\int_V d\mathbf{r}\sigma(\mathbf{r})\int d\Omega\psi^2(\mathbf{r},\mathbf{\Omega}). \tag{9}$$

Remembering (6) we see that

$$I_1\geq 0. \tag{10}$$

Since $f(\mathbf{\Omega},\mathbf{\Omega}')$ is positive definite [from (4)], we know that

$$\int\int d\Omega d\Omega' f(\mathbf{\Omega},\mathbf{\Omega}')[\psi(\mathbf{r},\mathbf{\Omega})-\psi(\mathbf{r},\mathbf{\Omega}')]^2\geq 0. \tag{11}$$

This together with the conditions of (3a) and (3b) gives

$$\int\int d\Omega d\Omega'\psi(\mathbf{r},\mathbf{\Omega})f(\mathbf{\Omega},\mathbf{\Omega}')\psi(\mathbf{r},\mathbf{\Omega}')\leq\int d\Omega\psi^2(\mathbf{r},\mathbf{\Omega}). \tag{12}$$

Hence

$$I_2\leq\int d\mathbf{r}\sigma(\mathbf{r})[c(\mathbf{r})-1]\int d\Omega\psi^2(\mathbf{r},\mathbf{\Omega})\leq 0. \tag{13}$$

The inequalities of (10) and (13) are compatible with the equality of (7) only if

$$\psi(\mathbf{r},\mathbf{\Omega})\equiv 0 \quad \text{in } V. \tag{14}$$

Two remarks should be made:

(a) The assumptions of the theorem are probably stronger than absolutely necessary. However, they are satisfactory for the applications of interest here.

(b) In case the region V extends to infinity (such as the exterior of a sphere or a half-space) the uniqueness is to be understood as applied to functions vanishing sufficiently rapidly. (Physically this means there are to be no sources at infinity.)

As a direct application of this theorem we note that to obtain the most general solution of (2) we need determine only two functions for a given region. These are $\psi_p(\mathbf{r},\mathbf{\Omega};\mathbf{r}_0,\mathbf{\Omega}_0)$ and $\psi_s(\mathbf{r},\mathbf{\Omega};\mathbf{r}_s,\mathbf{\Omega}_s)$. Here ψ_p satisfies Eq. (2) with $q(\mathbf{r},\mathbf{\Omega})=\delta(\mathbf{r}-\mathbf{r}_0)\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_0)$ and describes zero incoming flux. ($\delta(\mathbf{r})$ is the Dirac delta function: $\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_0)=0$, $\mathbf{\Omega}\cdot\mathbf{\Omega}_0\neq 1$ and $\int\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_0)d\Omega=1$.) ψ_s satisfies the homogeneous form of (2) with the condition

$$\psi_s(\mathbf{r},\mathbf{\Omega};\mathbf{r}_s,\mathbf{\Omega}_s)|_{\mathbf{r}\text{ on }S}=\delta_s(\mathbf{r}-\mathbf{r}_s)\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_s). \tag{15}$$

Here δ_s is the delta function on the surface S and $\mathbf{\Omega}_s$ is a direction such that $\mathbf{n}_0.\ \mathbf{\Omega}_s|_{\mathbf{r}=\mathbf{r}_s}<0$.

The solution of (2) with arbitrary source $q(\mathbf{r},\mathbf{\Omega})$ and arbitrary incident flux $\psi_{\text{inc}}(\mathbf{r}_s,\mathbf{\Omega}_s)$ is clearly

$$\psi(\mathbf{r},\mathbf{\Omega})=\int_S dS\int_{\mathbf{n}_0\cdot\mathbf{\Omega}_s<0}d\Omega_s\psi_s(\mathbf{r},\mathbf{\Omega};\mathbf{r}_s,\mathbf{\Omega}_s)\psi_{\text{inc}}(\mathbf{r}_s,\mathbf{\Omega}_s)$$
$$+\int_V d\mathbf{r}_0\int d\Omega_0\psi_p(\mathbf{r},\mathbf{\Omega};\mathbf{r}_0,\mathbf{\Omega}_0)q(\mathbf{r}_0,\mathbf{\Omega}_0). \tag{16}$$

Proof that this is the solution is obtained merely by noting that (a) the equation is satisfied; (b) the boundary conditions are satisfied; (c) the solution is unique.

In a sense to be shown below only ψ_s is necessary (ψ_p will be obtained from ψ_s).

It is we hope clear how uniqueness can be proved similarly for the more general situation envisaged in (1).

III. FUNDAMENTAL IDENTITY

Let $\psi_1(\mathbf{r},\mathbf{\Omega})$ be the solution of

$$\mathbf{\Omega}\cdot\nabla\psi_1(\mathbf{r},\mathbf{\Omega})+\sigma(\mathbf{r})\psi_1$$
$$=c(\mathbf{r})\sigma(\mathbf{r})\int f(\mathbf{\Omega},\mathbf{\Omega}')\psi_1(\mathbf{r},\mathbf{\Omega}')d\Omega'+q_1(\mathbf{r},\mathbf{\Omega}) \tag{17}$$

within a region V subject to $\psi_1(\mathbf{r},\mathbf{\Omega})=\psi_{1\,\text{inc}}(\mathbf{r},\mathbf{\Omega})$ (for $\mathbf{n}_0\cdot\mathbf{\Omega}<0$) on the bounding surface S. Here ψ_{inc} is a given function. Similarly we consider a solution ψ_2 corresponding to a source q_2 and a given incident flux $\psi_{2\,\text{inc}}$. In addition to ψ_2 we consider a function $\tilde{\Psi}_2(\mathbf{r},\mathbf{\Omega})$

which is the solution of

$$\Omega\cdot\nabla\tilde{\psi}_2(\mathbf{r},\Omega)+\sigma(\mathbf{r})\tilde{\psi}_2$$
$$=c(\mathbf{r})\sigma(\mathbf{r})\int f(-\Omega',-\Omega)\tilde{\psi}_2(\mathbf{r},\Omega')d\Omega'+q_2(\mathbf{r},\Omega),\quad (18a)$$

subject to

$$\tilde{\psi}_2(\mathbf{r},\Omega)=\tilde{\psi}_{2\ \mathrm{inc}}(\mathbf{r},\Omega)\equiv\psi_{2\ \mathrm{inc}}(\mathbf{r},\Omega)\qquad(\text{for } \mathbf{n}_0\cdot\Omega<0 \text{ on } S).$$

(The $\sim$ is associated with the indicated transposition operation on the scattering function f.) Let us substitute $-\Omega$ for Ω in (18a). This becomes

$$-\Omega\cdot\nabla\tilde{\psi}_2(\mathbf{r},-\Omega)+\sigma(\mathbf{r})\tilde{\psi}_2(\mathbf{r},-\Omega)$$
$$=c(\mathbf{r})\sigma(\mathbf{r})\int f(\Omega',\Omega)\tilde{\psi}_2(\mathbf{r},-\Omega')d\Omega'+q_2(\mathbf{r},-\Omega).\quad (18b)$$

Multiply (17) by $\tilde{\psi}_2(\mathbf{r},-\Omega)$ and (18b) by $\psi_1(\mathbf{r},\Omega)$. Subtract the second from the first and integrate over V and Ω. Applying Gauss' theorem we obtain

$$\int_S dS\int d\Omega\mathbf{n}_0\cdot\Omega\tilde{\psi}_2(\mathbf{r},-\Omega)\psi_1(\mathbf{r},\Omega)$$
$$=\int_V d\mathbf{r}c(\mathbf{r})\sigma(\mathbf{r})\int\int d\Omega d\Omega'\{\tilde{\psi}_2(\mathbf{r},-\Omega)f(\Omega,\Omega')$$
$$\times\psi_1(\mathbf{r},\Omega')-\psi_1(\mathbf{r},\Omega)f(\Omega',\Omega)\tilde{\psi}_2(\mathbf{r},-\Omega')\}$$
$$+\int_V d\mathbf{r}\int d\Omega\{\tilde{\psi}_2(\mathbf{r},-\Omega)q_1(\mathbf{r},\Omega)$$
$$-\psi_1(\mathbf{r},\Omega)q_2(\mathbf{r},-\Omega)\}.\quad (19)$$

Interchanging the names of the integration variables Ω and Ω' shows that the first term on the right in Eq. (19) vanishes identically. Since $\psi_1(\mathbf{r},\Omega)$ and $\tilde{\psi}_2(\mathbf{r},\Omega)$ are prescribed on S for $\mathbf{n}_0\cdot\Omega<0$, it is useful to rewrite Eq. (19) as

$$\int_S dS\int_{\mathbf{n}_0\cdot\Omega<0} d\Omega|\mathbf{n}_0\cdot\Omega|\{\tilde{\psi}_{2\ \mathrm{inc}}(\mathbf{r},\Omega)\psi_1(\mathbf{r},-\Omega)$$
$$-\psi_{1\ \mathrm{inc}}(\mathbf{r},\Omega)\tilde{\psi}_2(\mathbf{r},-\Omega)\}$$
$$=\int_V d\mathbf{r}\int d\Omega\{\tilde{\psi}_2(\mathbf{r},-\Omega)q_1(\mathbf{r},\Omega)$$
$$-\psi_1(\mathbf{r},\Omega)q_2(\mathbf{r},-\Omega)\}.\quad (20)$$

From this identity the results which have been called reciprocity principles follow as special cases. We give three examples here. In all these we suppose $f(\Omega,\Omega')=f(-\Omega',-\Omega)$, i.e., the scattering function has time reflection symmetry. From the uniqueness theorem we conclude that $\tilde{\psi}_2\equiv\psi_2$.

(a) Let ψ_1 correspond to a directional point source in direction Ω_1 at $\mathbf{r}_1$. ($q_1(\mathbf{r},\Omega)=\delta_2(\Omega\cdot\Omega_1)\delta(\mathbf{r}-\mathbf{r}_1)$, $\psi_{1\ \mathrm{inc}}=0$.) We denote this solution by $\psi_p(\mathbf{r},\Omega;\mathbf{r}_1,\Omega_1)$. Similarly let $\psi_2(\mathbf{r},\Omega)=\psi_p(\mathbf{r},\Omega;\mathbf{r}_2,\Omega_2)$. Inserting in Eq. (20) gives

$$\psi_p(\mathbf{r}_1,-\Omega_1;\mathbf{r}_2,\Omega_2)=\psi_p(\mathbf{r}_2,-\Omega_2;\mathbf{r}_1,\Omega_1).\quad (21)$$

Thus the angular density at $\mathbf{r}_1$, direction $-\Omega_1$, resulting from a unit point source in direction Ω_2 at $\mathbf{r}_2$, is equal to the angular density at $\mathbf{r}_2$, in direction $-\Omega_2$, due to a unit point source in direction Ω_1 at $\mathbf{r}_1$. From this many additional results follow. For example, on integrating the two sides of (21) over Ω_1 and Ω_2 we find the theorem: The density at $\mathbf{r}_1$ due to a unit isotropic source at $\mathbf{r}_2$ is equal to the density at $\mathbf{r}_2$ due to a unit isotropic source at $\mathbf{r}_1$.[2] Similarly on multiplying Eq. (21) by $\mathbf{w}\cdot\Omega$ ($\mathbf{w}$ an arbitrary unit vector) and then integrating over Ω_1 and Ω_2 we see that the component of the *current* in direction $\mathbf{w}$ at $\mathbf{r}_1$ due to an isotropic source at $\mathbf{r}_2$ is equal to the *density* at $\mathbf{r}_2$ due to the directional source $\mathbf{w}\cdot\Omega$ at $\mathbf{r}_1$.

(b) Let $\psi_1=\psi_s(\mathbf{r},\Omega;\mathbf{r}_{1s},\Omega_{1s})$, (i.e., $q_1=0$, $\psi_{1\ \mathrm{inc}}=\delta_s(\mathbf{r}_1-\mathbf{r}_{1s})\delta_2(\Omega\cdot\Omega_1)$), and let $\psi_2=\psi_s(\mathbf{r},\Omega;\mathbf{r}_{2s},\Omega_2)$. (Note that $\mathbf{n}_0\cdot\Omega_1|\mathbf{r}_{1s}<0$ and $\mathbf{n}_0\cdot\Omega_2|\mathbf{r}_{2s}<0$ is implied.) From Eq. (20) we find

$$|\mathbf{n}_2\cdot\Omega_2|\psi_s(\mathbf{r}_{2s},-\Omega_2;\mathbf{r}_{1s},\Omega_1)$$
$$=|\mathbf{n}_1\cdot\Omega_1|\psi(\mathbf{r}_{1s},-\Omega_1;\mathbf{r}_{2s},\Omega_2).\quad (22)$$

This says that the *emergent* angular distribution in direction $-\Omega_2$ at $\mathbf{r}_{2s}$ due to an *incident* beam at $\mathbf{r}_{1s}$ in direction Ω_1 times the absolute value of the cosine of the angle between Ω_2 and the normal to S at $\mathbf{r}_{2s}$ is equal to the emergent distribution at $\mathbf{r}_{1s}$ in direction $-\Omega_1$ due to an incident beam at $\mathbf{r}_{2s}$ in direction Ω_2 times the absolute value of the cosine of the angle between Ω_1 and the normal at $\mathbf{r}_{1s}$.

(c) Let $\psi_1=\psi_p(\mathbf{r},\Omega;\mathbf{r}_1,\Omega_1)$, $\psi_2=\psi_s(\mathbf{r},\Omega;\mathbf{r}_{2s},\Omega_s)$. Substituting in Eq. (20) we find

$$\psi_p(\mathbf{r}_{2s},-\Omega_2;\mathbf{r}_1,\Omega_1)=\frac{1}{|\mathbf{n}_2\cdot\Omega_2|}\psi_s(\mathbf{r}_1,-\Omega_1;\mathbf{r}_{2s},\Omega_2).\quad (23)$$

Thus, the emergent angular distribution in direction $-\Omega_2$ at $\mathbf{r}_{2s}$ due to a point source of direction Ω_1 at $\mathbf{r}_1$ is the product of the reciprocal of the absolute value of the cosine of the angle between the normal at $\mathbf{r}_{2s}$ and Ω_2 and the angular density at $\mathbf{r}_1$, direction $-\Omega_1$ due to an incident beam in direction Ω_2 at $\mathbf{r}_{2s}$. Hence, if we are primarily interested in emerging angular distributions *only* the function ψ_s is necessary.

A particular consequence of Eq. (23) that will be used below is obtained by integrating both sides over all Ω_1. This yields: the emergent angular distribution at $\mathbf{r}_{2s}$, direction $-\Omega_2$ is $1/(4\pi|\mathbf{n}_2\cdot\Omega_2|)$ times the

[2] By "density" and "current" we mean $\int\psi(\mathbf{r},\Omega)d\Omega$ and $\int\Omega\psi(\mathbf{r},\Omega)d\Omega$, respectively.

density at $\mathbf{r}_1$ due to an incident beam in direction $\mathbf{\Omega}_2$ at $\mathbf{r}_{2s}$.

IV. PROBLEMS WITH PLANE SYMMETRY

In these problems Eq. (2) simplifies (on introducing the optical thickness[3]) to

$$\mu\frac{\partial\psi(z,\mathbf{\Omega})}{\partial z}+\psi(z,\mathbf{\Omega})=c(z)\int f(\mathbf{\Omega},\mathbf{\Omega}')\psi(z,\mathbf{\Omega}')d\Omega'+q(z,\mathbf{\Omega}), \quad (24)$$

where μ is the cosine of the angle between $\mathbf{\Omega}$ and the positive z direction. The equation for $\tilde{\psi}$ is the same with $f(\mathbf{\Omega},\mathbf{\Omega}')$ replaced by $f(-\mathbf{\Omega}',-\mathbf{\Omega})$.

We can, of course, appropriately specialize Eq. (20) for this symmetry. It is, however, convenient to rederive the result. Since incident and emergent directions will be characterized by $\mu>0$ or $\mu<0$ it is suggested by the uniqueness theorem to make the decomposition (used by Chandrasekhar[1]) of ψ into $\psi_\pm$ defined by

$$\psi_\pm(z,\mathbf{\Omega})=\psi(z,\pm\mathbf{\Omega}), \quad (\mu>0). \quad (25)$$

Similarly we decompose our sources into $q_\pm(z,\mathbf{\Omega})$. For a given source q_1 Eq. (24) becomes the two equations

$$\pm\mu\frac{\partial\psi_\pm^{(1)}(z,\mathbf{\Omega})}{\partial z}+\psi_\pm^{(1)}=c(z)\int\{f(\pm\mathbf{\Omega},\mathbf{\Omega}')\psi_+^{(1)}(z,\mathbf{\Omega}') +f(\pm\mathbf{\Omega},-\mathbf{\Omega}')\psi_-^{(1)}(z,\mathbf{\Omega}')\}d\Omega'+q_\pm^{(1)}(z,\mathbf{\Omega}). \quad (26a, b)$$

(Here all $\mathbf{\Omega}$ are restricted to $\mu>0$.)

For the functions $\tilde{\psi}_\pm^{(2)}$ corresponding to a different source $q^{(2)}$ we have

$$\pm\mu\frac{\partial\tilde{\psi}_\pm^{(2)}}{\partial z}(z,\mathbf{\Omega})+\tilde{\psi}_\pm^{(2)}=c(z)\int\{f(-\mathbf{\Omega}',\mp\mathbf{\Omega})\tilde{\psi}_+^{(2)}(z,\mathbf{\Omega}') +f(\mathbf{\Omega}',\mp\mathbf{\Omega})\tilde{\psi}_-^{(2)}(z,\mathbf{\Omega}')\}d\Omega'+q_\pm^{(2)}(z,\mathbf{\Omega}). \quad (27a, b)$$

Multiply Eq. (26a) by $\tilde{\psi}_-^{(2)}(z,\mathbf{\Omega})$, Eq. (26b) by $\tilde{\psi}_+^{(2)}(z,\mathbf{\Omega})$, Eq. (27a) by $\psi_-^{(1)}(z,\mathbf{\Omega})$, and Eq. (27b) by $\psi_+^{(1)}(z,\mathbf{\Omega})$. Subtract the sum of the resulting second two equations from the sum of the resulting first two. Integrating $\mathbf{\Omega}$ over all angles with $\mu>0$ and z between a and b $(a>b)$ we find

$$\int d\Omega\mu\{\tilde{\psi}_-^{(2)}(z,\mathbf{\Omega})\psi_+^{(1)}(z,\mathbf{\Omega}) -\tilde{\psi}_+^{(2)}(z,\mathbf{\Omega})\psi_-^{(1)}(z,\mathbf{\Omega})\}_a{}^b$$
$$=\int_a^b dz\int d\Omega\{\tilde{\psi}_-^{(2)}(z,\mathbf{\Omega})q_+^{(1)}(z,\mathbf{\Omega}) +\tilde{\psi}_+^{(2)}(z,\mathbf{\Omega})q_-^{(1)}(z,\mathbf{\Omega})-\psi_-^{(1)}(z,\mathbf{\Omega})q_+^{(2)}(z,\mathbf{\Omega}) -\psi_+^{(1)}(z,\mathbf{\Omega})q_-^{(2)}(z,\mathbf{\Omega})\}. \quad (28)$$

Specializing to various sources and incident angular distributions the analogs of the theorems in Sec. III are obtained. It will be seen, though, that even more information is contained in these equations.

V. HALF-SPACE PROBLEMS—BASIC EQUATIONS

The fundamental problem of this type is: find functions $(\psi_\pm(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger))$ which satisfy Eqs. (26a, b) (with $q_\pm=0$) for $z\geq\bar{z}$ such that $\psi_+(\bar{z},\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)=\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger)$.

A convenient reformulation of this problem is the following: Find functions satisfying Eqs. (26a, b) everywhere which agree with the above $\psi_\pm(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)$ for $z>\bar{z}$ and which vanish for $z<\bar{z}$. Clearly this can only be achieved by having plane sources at $z=\bar{z}$. To see what these sources must be we can write our equations as

$$\pm\mu\frac{\partial\psi_\pm}{\partial z}(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)+\psi_\pm=c(z)\int\{\ \ \}+\lambda_\pm(\mathbf{\Omega})\delta(z-\bar{z}).$$

Integrating these equations from $\bar{z}-\epsilon$ to $\bar{z}+\epsilon$ (ϵ is infinitesimal) gives

$$\pm\mu\{\psi_\pm(\bar{z}+\epsilon,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)-\psi_\pm(\bar{z}-\epsilon,\mathbf{\Omega};\bar{z},\mathbf{e}^\dagger)\}=\lambda_\pm(\mathbf{\Omega}). \quad (29a, b)$$

Since

$$\psi_\pm(\bar{z}-\epsilon,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)=0$$

and

$$\psi_+(\bar{z}+\epsilon,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)=\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger),$$

we have

$$\lambda_+(\mathbf{\Omega})=\mu\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger), \quad (30a)$$

$$\lambda_-(\mathbf{\Omega})=-\mu\psi_-(\bar{z},\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger). \quad (30b)$$

Let us now apply Eq. (28) to $\psi_\pm(z,\mathbf{\Omega};z_1,\mathbf{\Omega}_1)$ and $\tilde{\psi}_\pm(z,\mathbf{\Omega};z_2,\mathbf{\Omega}_2)$ where $z_2>z_1$. The sources are

$$q_+^{(1,2)}(z,\mathbf{\Omega})=\mu\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_{1,2})\delta(z-z_{1,2}),$$
$$q_-^{(1)}(z,\mathbf{\Omega})=-\mu\psi_-(z_1,\mathbf{\Omega};z_1,\mathbf{e}_1)\delta(z-z_1), \quad (31)$$
$$q_-^{(2)}(z,\mathbf{\Omega})=-\mu\tilde{\psi}_-(z_2,\mathbf{\Omega};z_2,\mathbf{\Omega}_2)\delta(z-z_2).$$

---
[3] For example, see Case, deHoffmann, and Placzek, *Introduction to the Theory of Neutron Diffusion* (U. S. Government Printing Office, Washington, D. C., 1953), Vol. 1.

If we remember that $\tilde{\psi}_{\pm}(z,\mathbf{\Omega};z_2,\mathbf{\Omega}_2)=0$ for $z<z_2$ and require the functions to vanish as $z\to\infty$ we obtain

$$\mu_2\psi_-(z_2,\mathbf{\Omega}_2;z_1,\mathbf{\Omega}_1)=\int d\mathbf{\Omega}\mu\tilde{\psi}_-(z_2,\mathbf{\Omega};z_2,\mathbf{\Omega}_2)\psi_+(z_2,\mathbf{\Omega};z_1,\mathbf{\Omega}_1). \quad (32)$$

Suppose $z_2\to z_1$. Then $\psi_+(z_2,\mathbf{\Omega};z_1,\mathbf{\Omega}_1)\to\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_1)$ and we find the relation

$$\mu_2\psi_-(z_1,\mathbf{\Omega}_2;z_1,\mathbf{\Omega}_1)=\mu_1\tilde{\psi}_-(z_1,\mathbf{\Omega}_1;z_1,\mathbf{\Omega}_2). \quad (33a)$$

If in particular the reversal symmetry holds $(f(\mathbf{\Omega},\mathbf{\Omega}')=f(-\mathbf{\Omega}',-\mathbf{\Omega}))$, we have $\psi=\tilde{\psi}$ and Eq. (33a) becomes

$$\mu_2\psi_-(z_1,\mathbf{\Omega}_2;z_1,\mathbf{\Omega}_1)=\mu_1\psi_-(z_1,\mathbf{\Omega}_1;z_1,\mathbf{\Omega}_2). \quad (33b)$$

Up to this point there has been absolutely no use of "The Principle of Invariance." Equation (32) is generally valid. However, as Chandrasekhar[1] has so fully shown, considerable further progress is possible if we can call on this invariance. In the present context this means *only* the following theorem.

Theorem.—If $c(z)$ is independent of z,

$$\psi_{\pm}(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^{\dagger})=\psi_{\pm}(z-\bar{z},\mathbf{\Omega};0,\mathbf{\Omega}^{\dagger}).$$

Proof.—$\psi_{\pm}(z-\bar{z},\mathbf{\Omega};0,\mathbf{\Omega}^{\dagger})$ satisfy exactly the same equations and boundary conditions as $\psi_{\pm}(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^{\dagger})$. The uniqueness theorem then tells us they are identical.

Define S and $\tilde{S}$ by

$$S(\mathbf{\Omega},\mathbf{\Omega}')=\mu\psi_-(0,\mathbf{\Omega};0,\mathbf{\Omega}'),$$
$$\tilde{S}(\mathbf{\Omega},\mathbf{\Omega}')=\mu\psi_-(0,\mathbf{\Omega};0,\mathbf{\Omega}'),$$

and put $z_1=0$, $z_2=z$ in Eq. (32). We find

$$\mu\psi_-(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)=\int d\mathbf{\Omega}'\tilde{S}(\mathbf{\Omega}',\mathbf{\Omega})\psi_+(z,\mathbf{\Omega}';0,\mathbf{\Omega}_1). \quad (34a)$$

From Eq. (34a) complete knowledge of S and $\tilde{S}$ can be obtained. For example, if we put $z=0$ this becomes

$$S(\mathbf{\Omega},\mathbf{\Omega}_1)=\tilde{S}(\mathbf{\Omega}_1,\mathbf{\Omega}), \quad (35a)$$

which can be used to rewrite Eq. (34a) as

$$\mu\psi_-(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)=\int d\mathbf{\Omega}'S(\mathbf{\Omega},\mathbf{\Omega}')\psi_+(z,\mathbf{\Omega}';0,\mathbf{\Omega}_1). \quad (34b)$$

[It is convenient that all "reversed" functions have disappeared from the fundamental Eq. (34b) even though no symmetry assumptions on the scattering law have been made.] We note that Eq. (34b) is just Chandrasekhar's starting point for treating half-space problems.

Further information can be obtained from Eq. (34b) by considering the limit $z\to\infty$. Since in this limit we are far from the boundary $z=0$ it is to be expected that ψ will tend towards an infinite medium solution of the homogeneous Eq. (26a, b). Translational invariance implies such a solution in as exponential function of z times a function of $\mathbf{\Omega}$. The requirement $\psi\to 0$ as $z\to\infty$ shows the argument of the exponential to be negative (i.e., $\psi_{\pm}\to\phi_{\pm}(\mathbf{\Omega})e^{-\kappa z}$ where $\kappa\geq 0$ and $\phi_{\pm}(\mathbf{\Omega})e^{-\kappa z}$ is a solution of the homogeneous transfer equation). Substituting in Eq. (34b) we find that

$$\mu\phi_-(\mathbf{\Omega})=\int d\mathbf{\Omega}'S(\mathbf{\Omega},\mathbf{\Omega}')\phi_+(\mathbf{\Omega}'). \quad (35b)$$

The remaining information needed can be obtained by following Chandrasekhar's procedure and differentiating Eq. (34b) with respect to z. The transfer equations [Eq. (26a, b)] can then be used to eliminate $\partial\psi_{\pm}/\partial z$. Taking the limit $z\to 0$ and remembering that

$$\lim_{z\to 0}\psi_+(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)=\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_1),$$

and

$$\lim_{z\to 0}\psi_-(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)=S(\mathbf{\Omega},\mathbf{\Omega}_1)/\mu,$$

we obtain

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu_1}\right)S(\mathbf{\Omega}_0,\mathbf{\Omega}_1)=c\int\int d\mathbf{\Omega}'d\mathbf{\Omega}''f(\mathbf{\Omega}',-\mathbf{\Omega}'')\times\left[\delta_2(-\mathbf{\Omega}_0\cdot\mathbf{\Omega}')+\frac{S(\mathbf{\Omega}_0,\mathbf{\Omega}')}{\mu'}\right]\times\left[\delta_2(-\mathbf{\Omega}''\cdot\mathbf{\Omega}_1)+\frac{S(\mathbf{\Omega}'',\mathbf{\Omega}_1)}{\mu''}\right]. \quad (36)$$

Here, for compactness in writing, we use the conventions that integrals are over *all* directions and $S(\mathbf{\Omega},\mathbf{\Omega}')$ is taken zero for μ or $\mu'<0$.

Thus, using only the reciprocity identity and the translational invariance the nonlinear integral equation [Eq. (36) and the condition on the solution Eq. (35b)] are obtained. Chandrasekhar[1] has shown in many cases that these uniquely determine S.[4] Given S the determination of $\psi_{\pm}(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)$ is well known and straightforward. Examples are given below.

For further applications of the reciprocity relations it is useful to specialize f. Let us assume, as is true in many cases of interest, that

$$f(\mathbf{\Omega}',\mathbf{\Omega}'')=\Phi_1(\mu')\Phi_2(\mu'')/4\pi,$$

where Φ_1, Φ_2 are even functions. (The physical origin of this even requirement will be seen later.) Then Eq. (36) becomes

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu_1}\right)S(\mathbf{\Omega}_0,\mathbf{\Omega}_1)=\frac{c}{4\pi}\mathcal{H}(\mu_0)\mathcal{G}(\mu_1), \quad (37)$$

[4] A general proof is possible. Since it introduces many concepts foreign to the present discussion and is applicable to a wider class of equations it will be presented elsewhere.

where

$$\mathcal{H}(\mu_0)=\Phi_1(\mu_0)+\int \frac{S(\Omega_0,\Omega')}{\mu'}\Phi_1(\mu')d\Omega', \quad (38a)$$

and

$$\mathcal{G}(\mu_1)=\Phi_2(\mu_1)+\int \frac{S(\Omega'',\Omega_1)}{\mu''}\Phi_2(\mu'')d\Omega''. \quad (38b)$$

Substituting for S from Eq. (37) gives the coupled integral equations

$$\mathcal{H}(\mu_0)=\Phi_1(\mu_0)+\frac{c}{2}\mu_0\mathcal{H}(\mu_0)\int_0^1 \frac{\Phi_1(\mu')\mathcal{G}(\mu')d\mu'}{\mu_0+\mu'}, \quad (39a)$$

$$\mathcal{G}(\mu_1)=\Phi_2(\mu_1)+\frac{c}{2}\mu_1\mathcal{G}(\mu_1)\int_0^1 \frac{\Phi_2(\mu'')\mathcal{H}(\mu'')d\mu''}{\mu_1+\mu''}. \quad (39b)$$

These are readily reducible to a single integral equation. Let $\mathcal{H}(\mu)=\Phi_1(\mu)H(\mu)$, $\mathcal{G}(\mu)=\Phi_2(\mu)H(\mu)$. The coupled equations are then equivalent to

$$H(\mu)=1+\frac{c}{2}\mu H(\mu)\int_0^1 \frac{\Psi(\mu')H(\mu')d\mu'}{\mu+\mu'}, \quad (40)$$

where

$$\Psi(\mu')=\Phi_1(\mu')\Phi_2(\mu'). \quad (41)$$

The equation for H is thus just in the standard form discussed in detail by Chandrasekhar.[1]

For S we have

$$S(\Omega_0,\Omega_1)=\frac{c}{4\pi}\frac{\mu_0\mu_1}{\mu_0+\mu_1}\Phi_1(\mu_0)\Phi_2(\mu_1)H(\mu_0)H(\mu_1), \quad (42a)$$

while

$$\tilde{S}(\Omega_0,\Omega_1)=\frac{c}{4\pi}\frac{\mu_0\mu_1}{\mu_0+\mu_1}\Phi_2(\mu_0)\Phi_1(\mu_1)H(\mu_0)H(\mu_1). \quad (42b)$$

VI. HALF-SPACE GREEN'S FUNCTION

For simplicity we restrict ourselves in the following example to isotropic scattering (i.e., $\Phi_1=\Phi_2=1$). Clearly[5] the discussion can be restricted to isotropic sources. The problem is then to find the solution $(\psi_{p\pm}(z,\Omega;z_0))$ of

$$\pm\mu\frac{\partial\psi_{p\pm}}{\partial z}(z,\Omega;z_0)+\psi_{p\pm}$$

$$=\frac{c}{4\pi}\int d\Omega'\{\psi_{p+}(z,\Omega')+\psi_{p-}(z,\Omega')\}d\Omega'+\delta(z-z_0)/4\pi, \quad (43)$$

subject to

$$\lim_{z\to 0}\psi_{p+}(z,\Omega;z_0)=0, \quad (44)$$

and

$$\lim_{z\to\infty}\psi_{p\pm}(z,\Omega;z_0)=0.$$

The remark at the end of Sec. III states that

$$\psi_{p-}(0,\Omega;z_0)=\frac{1}{4\pi\mu}\rho(z_0;0,\mu), \quad (45)$$

where $\rho(z_0;0,\mu)=\int d\Omega'\psi(z_0,\Omega';0,\Omega)$ is the density at z_0 due to an incident beam in direction Ω on the plane $z=0$. ρ is obtained in the following well-known way. Formal integration of Eqs. (26a, b) yields

$$\psi_+(z,\Omega)=e^{-z/\mu}\left\{\psi_+(0,\Omega)+\frac{1}{\mu}\int_0^z e^{+z'/\mu}Q_+(z',\Omega)dz'\right\}, \quad (46a)$$

and

$$\psi_-(z,\Omega)=e^{z/\mu}\left\{\frac{1}{\mu}\int_z^\infty e^{-z'/\mu}Q_-(z',\Omega)dz'\right\}, \quad (46b)$$

where

$$Q_\pm(z,\Omega)=c(z)\int\{f(\pm\Omega,\,\Omega')\psi_+(z,\Omega')$$
$$+f(\pm\Omega,\,-\Omega')\psi_-(z,\Omega')\}d\Omega'+q_\pm(z,\Omega). \quad (47)$$

In the present instance of isotropic scattering, constant c, and isotropic sources $q_0(z)/4\pi$ this says

$$\mu\psi_-(0,\mu)=\int_0^\infty e^{-z'/\mu}\left\{\frac{c}{4\pi}\rho(z')+\frac{q_0(z')}{4\pi}\right\}dz', \quad (48)$$

or

$$\frac{1}{p}\psi_-(0,1/p)=\text{ Laplace transform of }\left\{\frac{c}{4\pi}\rho+\frac{q_0}{4\pi}\right\}. \quad (49)$$

Applying the Fourier-Mellin inversion theorem gives

$$\frac{c}{4\pi}\rho(z_0;0,\mu)=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{1}{p}\psi_-(0,1/p;0,\mu)e^{pz_0}dp. \quad (50)$$

In Sec. V it was found that

$$\frac{1}{p}\psi_-(0,1/p;0,\mu)=\frac{c}{4\pi}\frac{H(\mu)H(1/p)}{p+1/\mu}, \quad (51)$$

where

$$H(u)=1+\frac{c}{2}\mu H(\mu)\int_0^1\frac{H(\mu')d\mu'}{\mu+\mu'}. \quad (52)$$

Therefore,

$$\rho(z_0;0,\mu)=\frac{H(\mu)}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{H(1/p)e^{pz_0}}{p+1/\mu}dp. \quad (53)$$

The work of Chandrasekhar[1] and others[3] enables us to regard $H(\mu)$ $(0\leq\mu\leq 1)$ as a known function. ("Known" here means in the sense of numerical tables. While it is true that analytical formulas are available these are so complicated as to be virtually useless.) The main problem is to express the contour integral in terms of "known" functions. This can be done with the aid of

[5] For example, see reference 3, p. 122.

the condition given by Eq. (35) which has so far been ignored. The infinite medium form of the homogeneous equations (26a, b) for isotropic scattering and constant c have the unique solution which vanishes for $z\to\infty$

$$\psi_\pm(z,\Omega)=(\text{constant})[1/(1\mp\kappa\mu)]e^{-\kappa z},$$

where κ is the positive root of the equation

$$1-c(\tanh^{-1}\kappa)/\kappa=0.$$

Hence in Eq. (35b) we have

$$\phi_\pm(\Omega)=1/(1\mp\kappa\mu). \tag{54}$$

Substituting the expression for S in terms of the H function we find from Eq. (35b) that

$$\int_0^1 \frac{\mu' H(\mu')d\mu'}{(\mu+\mu')(1-\kappa\mu')}=\frac{2}{c}\frac{1}{H(\mu)}\frac{1}{(1+\kappa\mu)}. \tag{55}$$

However, on decomposing into partial fractions we have the identity

$$\frac{\mu'}{(\mu+\mu')(1-\kappa\mu')}=\frac{1}{1+\kappa\mu}\left\{\frac{-\mu}{\mu+\mu'}+\frac{1}{\kappa}\frac{1}{(1/\kappa)-\mu'}\right\}. \tag{56}$$

On substituting this into Eq. (55) we obtain integrals readily evaluated on recalling the integral equation [Eq. (52)] satisfied by $H(\mu)$. The result is

$$\int_0^1 \frac{\mu' H(\mu')d\mu'}{(\mu+\mu')(1-\kappa\mu')}=\frac{2}{c}\frac{1}{1+\kappa\mu}\left\{\frac{1}{H(\mu)}-\frac{1}{H(-1/\kappa)}\right\}. \tag{57}$$

Comparing Eqs. (57) and (55) we see that $H(p)$ has a pole at $p=-1/\kappa$.

The integral in Eq. (53) can now be evaluated in the limit of large z_0. Deforming the contour to the left we see that the dominant term in an asymptotic expansion of $\rho(z_0;0,\mu)$ is given by the singularity of $H(1/p)$ furthest to the right in the complex plane. This is just the pole at $p=-\kappa$. Hence

$$\rho(z_0;0,\mu)\underset{z_0\to\infty}{\longrightarrow}(\text{constant})e^{-\kappa z_0}\frac{\mu H(\mu)}{1-\kappa\mu}. \tag{58}$$

The reciprocity relation (Eq. 45) says then that the emergent angular distribution from a half-space with a plane source far from the boundary is proportional to $H(\mu)/(1-\kappa\mu)$. Alternatively, $H(\mu)/(1-\kappa\mu)$ is the emergent angular distribution for the Milne problem (i.e., the problem of finding a solution of the homogeneous transfer equation for the half-space which is $\sim e^{\kappa z}$ for large z). Thus

$$\rho_m(z)=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{H(1/p)e^{pz}}{p-\kappa}dp. \tag{59}$$

This is a "known" function with thoroughly investigated properties.[3] (The normalization[6] chosen here is such that $\rho_m(0)=1$.)

Comparing Eqs. (53) and (59) it is apparent that

$$\left(\frac{d}{dz_0}+\frac{1}{\mu}\right)\frac{\rho(z_0;0,\mu)}{H(\mu)}=\eta_m(z_0), \tag{60}$$

where we define $\eta_m(z_0)$ by

$$\eta_m(z_0)=\left(\frac{d}{dz_0}-\kappa\right)\rho_m(z_0).$$

An elementary integration then gives

$$\rho(z_0;0,\mu)=H(\mu)\left\{e^{-z_0/\mu}+\int_0^{z_0}e^{-(z_0-z')/\mu}\eta_m(z')dz'\right\}. \tag{61}$$

The emergent angular distribution for a unit isotropic plane source at z_0 is given by Eq. (45) in conjunction with Eq. 61).

The density $\rho(z;z_0)$ resulting from an isotropic plane source is obtained by inverting the Laplace transform of Eq. (49). Thus

$$c\rho(z;z_0)+\delta(z-z_0)=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}e^{pz}H(1/p)\times\left\{e^{-pz_0}+\int_0^{z_0}e^{-p(z_0-z')}\eta_m(z')dz'\right\}dp. \tag{62}$$

The contour integral is readily evaluated since it is seen from Eq. (59) that

$$\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}H(1/p)e^{pz}dp=\delta(z-z_0)+S(z)\eta_m(z), \tag{63}$$

where

$$S(z)=0,\quad z<0$$
$$=1,\quad z>0.$$

Using this result in Eq. (62) it follows that

$$c\rho(z;z_0)=\eta_m(|z-z_0|)+\int_{|z-z_0|/2}^{(z+z_0)/2}\eta_m\left(z'+\frac{|z-z_0|}{2}\right)\times\eta_m\left(z'-\frac{|z-z_0|}{2}\right)dz'. \tag{64}$$ [7]

A more convenient form of this equation is one in which the corrections to the infinite medium plane

[6] The solution of the Milne problem given in reference 3 is $[\kappa/(1-c)^{\frac{1}{2}}]\rho_m(z)$.

[7] A result equivalent to this has been obtained by B. Davison and G. Placzek, Canadian Report MT-118.

source solution due to the boundary $z=0$ are given explicitly. This can be obtained as follows: Let z, $z_0 \to \infty$ while $|z-z_0|$ is kept finite. The result must certainly be the infinite medium plane source solution[8] $\rho_\infty(z-z_0)$ i.e.,

$$c\rho_\infty(z-z_0)=\eta_m(|z-z_0|)+\int_{|z-z_0|/2}^{\infty}\eta_m\left(z'+\frac{|z-z_0|}{2}\right)\times\eta_m\left(z'-\frac{|z-z_0|}{2}\right)dz'. \quad (65)$$

Therefore, on making a simplifying change of variable in the additional term in Eq. (64) we have

$$\rho(z;z_0)=\rho_\infty(z-z_0)-\frac{1}{c}\int_0^\infty \eta_m(\bar{z}+z)\eta_m(\bar{z}+z_0)d\bar{z}. \quad (66)$$

If desired the angular distributions can be obtained using Eqs. (66) and (46a, b).

VII. GREEN'S FUNCTION FOR ADJACENT HALF-SPACES

Generalizing the problem considered in the previous section we will here construct the Green's function for adjacent half-spaces. For simplicity we again restrict ourselves to isotropic scattering. The media for $z>0$ (region 1) and $z<0$ (region 2) are to be characterized by constants c_1 and c_2, respectively. The unit isotropic source will be at z_0 where $z_0>0$.

It is useful to start with the simpler problem of a unit directional plane source at $z=0$ emitting in direction Ω_1 into region 1 ($\mu_1>0$). The first example of a reciprocity relation discussed in Sec. III relates the density at z_0 due to this source to the emergent angular distribution at $z=0$ due to the unit isotropic source at z_0. Specifically

$$\psi_{p-}(0,\mathbf{\Omega}_1;z_0)=\rho(z_0;0,\mathbf{\Omega}_1)/4\pi. \quad (67)$$

Our program will be the following

(1) We will find the emergent distribution from region 1 due to the directional source at $z=0$.

(2) From this emergent distribution the density at z_0 due to the directional source will be obtained.

(3) Using (67) the emergent distribution from "1" due to the isotropic plane source is then known. From this we can determine:

(a) The density for $z>0$.
(b) The emergent distribution from "2."

(4) From the emergent distribution from "2" the density for $z<0$ due to the isotropic source will be found.

The first step is then to solve the equations

$$\pm\mu\frac{\partial}{\partial z}\psi_\pm(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)+\psi_\pm=\frac{c(z)}{4\pi}\int\{\psi_+(z,\mathbf{\Omega}')+\psi_-(z,\mathbf{\Omega}')\}d\mathbf{\Omega}'+\begin{cases}\delta(z)\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_1)\\ 0,\end{cases} \quad (68a, b)$$

where

$$c(z)=c_1,\quad z>0$$
$$=c_2,\quad z<0.$$

Denote the solutions for $z>0$ and $z<0$ by superscripts [1] and [2], respectively. Integrating Eqs. (68a, b) from slightly less than to slightly greater than zero gives

$$\pm\mu[\psi_\pm^{(1)}(0,\mathbf{\Omega})-\psi_\pm^{(2)}(0,\mathbf{\Omega})]=\begin{cases}\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}_1) & (69a)\\ 0. & (69b)\end{cases}$$

From the discussion in Sec. V we also know that

$$\psi_-^{(1)}(0,\mathbf{\Omega})=\int d\mathbf{\Omega}'\frac{S^{(1)}(\mathbf{\Omega},\mathbf{\Omega}')}{\mu}\psi_+^{(1)}(0,\mathbf{\Omega}'), \quad (70a)$$

and

$$\psi_+^{(2)}(0,\mathbf{\Omega})=\int d\mathbf{\Omega}'\frac{S^{(2)}(\mathbf{\Omega},\mathbf{\Omega}')}{\mu}\psi_-^{(2)}(0,\mathbf{\Omega}'), \quad (70b)$$

where $S^{(1)}$, $S^{(2)}$ are the S functions corresponding to c_1 and c_2.

Eliminating $\psi_+^{(1)}$ and $\psi_-^{(2)}$ from Eqs. (70a, b) by means of the relations of Eqs. (69a, b) we obtain the integral equations

$$\psi_-^{(1)}(0,\mathbf{\Omega})=\frac{c_1}{4\pi}\frac{H_1(\mu)H_1(\mu_1)}{\mu+\mu_1}+\frac{c_1}{2}H_1(\mu)\int_0^1\frac{\mu' H_1(\mu')}{\mu+\mu'}\psi_-^{(2)}(0,\mathbf{\Omega}')d\mu',$$
$$\psi_+^{(2)}(0,\mathbf{\Omega})=\frac{c_2}{2}H_2(\mu)\int_0^1\frac{\mu' H_2(\mu')}{\mu+\mu'}\psi_-^{(1)}(0,\mathbf{\Omega}')d\mu'. \quad (71)$$

Fortunately, these equations are readily solved. The identity

$$\frac{\mu'}{(\mu+\mu')(\mu'+a)}=\frac{1}{\mu-a}\left\{\frac{\mu}{\mu+\mu'}-\frac{a}{\mu'+a}\right\}$$

together with the integral equation satisfied by an H function gives

$$\int_0^1\frac{\mu' H(\mu')d\mu'}{(\mu+\mu')(\mu'+a)}=\frac{2}{c}\frac{1}{\mu-a}\left\{\frac{1}{H(a)}-\frac{1}{H(\mu)}\right\}. \quad (72)$$

[8] This is discussed in detail in reference 3.

This suggests that

$$\psi_-^{(1)}(0,\mathbf{\Omega})=\{\alpha_1+\beta_1H_1(\mu)/H_2(\mu)\}/(\mu+\mu_1),$$
$$\psi_+^{(2)}(0,\mathbf{\Omega})=\{\alpha_2+\beta_2H_2(\mu)/H_1(\mu)\}/(\mu-\mu_1), \quad (73)$$

where α_1, α_2, β_1, β_2 are constants to be determined.[9] Substituting Eq. (73) into Eq. (71) gives

$$\{\alpha_1+\beta_1H_1(\mu)/H_2(\mu)\}/(\mu+\mu_1)$$
$$=\frac{H_1(\mu)}{(\mu+\mu_1)}\left[\frac{c_1}{4\pi}H_1(\mu_1)+\frac{\alpha_2}{H_1(-\mu_1)}+\frac{\beta_2c_1/c_2}{H_2(-\mu_1)}\right]-\frac{\alpha_2}{(\mu+\mu_1)}-\frac{\beta_2H_1(\mu)c_1/c_2}{(\mu+\mu_1)H_2(\mu)},$$
$$\{\alpha_2+\beta_2H_2(\mu)/H_1(\mu)\}/(\mu-\mu_1)$$
$$=\frac{H_2(\mu)}{(\mu-\mu_1)}\left[\frac{\alpha_1}{H_2(\mu_1)}+\frac{\beta_1c_2/c_1}{H_1(\mu_1)}\right]-\frac{\alpha_1}{(\mu-\mu_1)}-\frac{\beta_1H_2(\mu)c_2/c_1}{(\mu-\mu_1)H_1(\mu)}. \quad (74)$$

These equations are compatible if

$$\alpha_1=-\alpha_2,\quad \beta_1=-\beta_2c_1/c_2,$$

$$\frac{c_1}{4\pi}H_1(\mu_1)+\frac{\alpha_2}{H_1(-\mu_1)}+\frac{\beta_2c_1/c_2}{H_2(-\mu_1)}=0,$$

and

$$\frac{\alpha_1}{H_2(\mu_1)}+\frac{\beta_1c_2/c_1}{H_1(\mu_1)}=0.$$

Solving for the constants we obtain

$$\beta_1=\frac{c_1}{4\pi}\frac{H_1(\mu_1)}{H_2(\mu_1)}\{\ \},\quad \beta_2=-\frac{c_2}{4\pi}\frac{H_1(\mu_1)}{H_2(\mu_1)}\{\ \},$$

$$\alpha_1=-\frac{c_2}{4\pi}\{\ \},\qquad \alpha_2=\frac{c_2}{4\pi}\{\ \},$$

where

$$\{\ \}=\left[\frac{1}{H_2(\mu_1)H_2(-\mu_1)}-\frac{c_2/c_1}{H_1(\mu_1)H_1(-\mu_1)}\right]^{-1}. \quad (75)$$

The expression of Eq. (75) simplifies on using the identity[10]

$$\frac{1}{H(z)H(-z)}=1-c(\tanh^{-1}z)/z, \quad (76)$$

which follows directly from the integral equation for the H functions. Thus we obtain

$$\psi_-^{(1)}=\frac{-c_1c_2}{4\pi(c_1-c_2)}\frac{1}{(\mu+\mu_1)}\left\{1-\frac{c_1H_1(\mu_1)H_1(\mu)}{c_2H_2(\mu_1)H_2(\mu)}\right\}, \quad (77a)$$

$$\psi_+^{(2)}=\frac{-c_1c_2}{4\pi(c_1-c_2)}\frac{1}{(\mu-\mu_1)}\left\{-1+\frac{H_1(\mu_1)H_2(\mu)}{H_2(\mu_1)H_1(\mu)}\right\}. \quad (77b)$$

Inverting Eq. (49) shows the neutron density at $z_0>0$ due to the directional source at $z=0$ is given by

$$\frac{c_1}{4\pi}\rho(z_0;0,\mathbf{\Omega}_1)=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{\psi_-^{(1)}(0,1/p;0,\mathbf{\Omega}_1)}{p}e^{pz_0}dp, \quad (78)$$

or

$$\rho(z_0;0,\mathbf{\Omega}_1)=J_1+J_2,$$

where

$$J_1=\frac{-c_2}{c_1-c_2}\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{e^{pz_0}dp}{1+p\mu_1}=\frac{-c_2}{c_1-c_2}\frac{e^{-z_0/\mu_1}}{\mu_1}, \quad (79)$$

and

$$J_2=\frac{c_1}{c_1-c_2}\frac{H_1(\mu_1)}{H_2(\mu_1)}\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{H_1(1/p)e^{pz_0}dp}{H_2(1/p)(1+p\mu_1)}. \quad (80)$$

As in Sec. VI, J_2 can be related to "known" functions. When $z_0\rightarrow\infty$ the principal contribution in Eq. (78) comes from the pole of $H_1(1/p)$ at $p=-\kappa_1$.

$$[1-c_1(\tanh^{-1}\kappa_1)/\kappa_1=0.]$$

Hence

$$\rho(z_0;0,\mathbf{\Omega}_1)\underset{z_0\rightarrow\infty}{\rightarrow}(\text{constant})\frac{e^{-\kappa_1z_0}}{1-\kappa_1\mu_1}\frac{H_1(\mu_1)}{H_2(\mu_1)}. \quad (81)$$

Thus, by Eq. (67), $H_1(\mu_1)/H_2(\mu_1)(1-\kappa_1\mu_1)$ is the emergent distribution from region "1" resulting from a source located far from the boundary. In other words, this is the emergent distribution for the generalized Milne problem—the problem of finding the solution of the homogeneous form of Eq. (68) subject to the condition of behaving as e^{κ_1z} for $z\rightarrow\infty$. Let us call the density for $z>0$ corresponding to this problem $\bar{\rho}_m^{(1)}(z)$. The inverse of the Laplace transformation gives

$$\bar{\rho}_m^{(1)}(z)=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{H_1(1/p)e^{pz}dp}{(p-\kappa_1)H_2(1/p)}. \quad (82)$$

This can be expressed in terms of the known functions $H_2(\mu)$ and $\rho_m^{(1)}(z)$. [Here $\rho_m^{(1)}(z)$ is defined as in Eq. (59) by

$$\rho_m^{(1)}(z)=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{H_1(1/p)}{p-\kappa_1}e^{pz}dp.]$$

Thus, on using the integral equation for H_2 in the form

$$\frac{1}{H_2(1/p)}=1-\frac{c_2}{2}\int_0^1\frac{H_2(\mu')d\mu'}{1+p\mu'} \quad (83)$$

[9] This form is also suggested by the solution obtained for a similar set of equations by S. Chandrasekhar, Can. J. Research A29, 14 (1951).

[10] See reference 1, p. 116.

it is found that

$$\bar{\rho}_m^{(1)}(z)=\rho_m^{(1)}(z)-\frac{c_2}{2}\int_0^z \rho_m^{(1)}(z')dz' \times\int_0^1 e^{-(z-z')/\mu'}\frac{H_2(\mu')d\mu'}{\mu'}. \quad (84)$$

In terms of $\bar{\eta}_m^{(1)}(z)\equiv(d/dz-\kappa_1)\bar{\rho}_m^{(1)}(z)$ we can write Eq. (80) as

$$J_2=\frac{c_1}{c_1-c_2}\frac{H_1(\mu_1)}{H_2(\mu_1)}\frac{1}{\mu_1} \times\left\{e^{-z_0/\mu_1}+\int_0^{z_0} e^{-(z_0-z')/\mu_1}\bar{\eta}_m^{(1)}(z')dz'\right\}. \quad (85)$$

Using Eq. (67) the emergent angular distribution from region "1" due to a unit isotropic plane source at $z_0>0$ is

$$\psi_{p-}(0,\mathbf{\Omega}_1;z_0)=\frac{-c_2}{4\pi(c_1-c_2)}\frac{e^{-z_0/\mu_1}}{\mu_1} +\frac{c_1}{4\pi(c_1-c_2)}\frac{H_1(\mu_1)}{H_2(\mu_1)}\frac{1}{\mu_1}\{\ \}, \quad (86)$$

where $\{\ \}$ is the same as in Eq. (85).

The inversion theorem gives for the density at $z>0$

$$\bar{\rho}(z;z_0)=\frac{1}{c_1-c_2}\bar{\eta}_m^{(1)}(|z-z_0|) +\frac{1}{c_1-c_2}\int_{|z-z_0|/2}^{(z+z_0)/2}\bar{\eta}_m^{(1)}\left(z'+\frac{|z-z_0|}{2}\right) \times\bar{\eta}_m^{(1)}\left(z'-\frac{|z-z_0|}{2}\right)dz'. \quad (87)$$

Again considerations of z, $z_0\to\infty$, $|z-z_0|$ finite, identifies the terms which describe a plane source in an infinite medium of properties "1" ($\rho_\infty^{(1)}(z-z_0)$) and the correction terms due to the boundary. The result is

$$\bar{\rho}(z;z_0)=\rho_\infty^{(1)}(z-z_0)-\frac{1}{c_1-c_2}\int_0^\infty\bar{\eta}_m^{(1)}(z'+z_0) \times\bar{\eta}_m^{(1)}(z'+z)dz'. \quad (z,z_0>0). \quad (88)$$

[It is immediately apparent that Eq. (88) reduces to Eq. (66) when $c_2\to0$ and to $\rho_\infty^{(1)}(z-z_0)$ when $c_2\to c_1$.]

The emergent angular distribution from region "2" due to the plane source at z_0 in region "1" is

$$\psi_{p+}(0,\mathbf{\Omega};z_0)=\int\frac{S^{(2)}(\mathbf{\Omega},\mathbf{\Omega}')}{\mu}\psi_{p-}(0,\mathbf{\Omega}';z_0)d\Omega'. \quad (89)$$

Expressing $S^{(2)}$ in terms of H_2, using the integral equation for H_2 and remembering that $z_0>0$ we find

$$\psi_{p+}(0,\mathbf{\Omega};z_0)=\frac{-c_2}{4\pi(c_1-c_2)}\frac{H_2(\mu)}{H_1(\mu)}\frac{1}{2\pi i} \times\int_{\beta-i\infty}^{\beta+i\infty}\frac{e^{pz_0}}{1-\mu p}\frac{H_1(1/p)}{H_2(1/p)}dp. \quad (90a)$$

With the aid of Eq. (82) it is seen that

$$\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{e^{pz_0}}{1-\mu p}\frac{H_1(1/p)}{H_2(1/p)}dp =\frac{e^{z_0/\mu}}{\mu}\int_{z_0}^\infty e^{-z'/\mu}\bar{\eta}_m^{(1)}(z')dz' =\frac{e^{z_0/\mu}}{\mu}\left\{\frac{H_1(\mu)}{H_2(\mu)}-\int_0^{z_0}e^{-z'/\mu}\bar{\eta}_m^{(1)}(z')dz'\right\}.$$

Hence

$$\psi_{p+}(0,\mathbf{\Omega};z_0)=\frac{-c_2}{4\pi(c_1-c_2)}\frac{1}{\mu}\left\{e^{z_0/\mu}-\frac{H_2(\mu)}{H_1(\mu)} \times\int_0^{z_0}e^{(z_0-z')/\mu}\bar{\eta}_m^{(1)}(z')dz'\right\}. \quad (90b)$$

The Fourier-Mellin inversion theorem then gives

$$\bar{\rho}(z;z_0)=\frac{-1}{c_1-c_2}\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}e^{-pz}\left\{e^{pz_0}-\frac{H_2(1/p)}{H_1(1/p)} \times\int_0^{z_0}e^{p(z_0-z')}\bar{\eta}_m^{(1)}(z')dz'\right\}, \quad (z<0,\ z_0>0). \quad (91)$$

Changing the order of integration yields

$$\bar{\rho}(z;z_0)=\frac{1}{c_1-c_2}\int_0^{z_0}\bar{\eta}_m^{(1)}(z')dz'\frac{1}{2\pi i} \times\int_{\beta-i\infty}^{\beta+i\infty}e^{p(z_0-z'-z)}\frac{H_2(1/p)}{H_1(1/p)}dp. \quad (92)$$

However, it is readily shown that

$$\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{H_2(1/p)}{H_1(1/p)}e^{-pz}dp=\bar{\eta}_m^{(2)}(z)\equiv\left(\frac{d}{dz}-\kappa_1\right)\bar{\rho}_m^{(2)}(z), \quad (93)$$

where $\bar{\rho}_m^{(2)}(z)$ is the density in region "2" in the generalized Milne problem. Expressed in terms of $\bar{\eta}_m^{(2)}$

Eq. (92) is

$$\bar{\rho}(z;z_0)=\frac{1}{c_1-c_2}\int_{(z-z_0)/2}^{(z+z_0)/2}\bar{\eta}_m{}^{(1)}\left(z'-\frac{z-z_0}{2}\right)\times\bar{\eta}_m{}^{(2)}\left(z'+\frac{z-z_0}{2}\right)dz',$$

where $z<0$, $z_0>0$. [Explicit formulas for $\bar{\rho}_m{}^{(2)}$, analogous to Eq. (84) for $\bar{\rho}_m{}^{(1)}$, in terms of $H_1(\mu)$ and the solution of the standard Milne problem for a medium of properties "2" ($\rho_m{}^{(2)}$) are readily obtained.]

VIII. EQUATIONS FOR A PLANE SLAB

Finally it will be shown that Chandrasekhar's equations[11] which determine the reflection and transmission of a finite slab follow directly from the identity of Eq. (28).

The problem is to determine the solution of the homogeneous form of Eq. (26) in a region $\bar{z}\leq z\leq\bar{z}+\tau$ subject to the boundary conditions

$$\psi_{\text{inc}}(\bar{z},\mathbf{\Omega})\equiv\psi_+(\bar{z},\mathbf{\Omega})=\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger),$$
$$\psi_{\text{inc}}(\bar{z}+\tau,\,\Omega)\equiv\psi_-(\bar{z}+\tau,\,\Omega)=0. \quad (94)$$

Leaving the τ dependence implicit, we denote the solution by $\Psi_\pm(z,\mathbf{\Omega},\bar{z},\mathbf{\Omega}^\dagger)$. As before we consider a solution which coincides with Ψ for $\bar{z}<z<\bar{z}+\tau$ and vanishes outside this region. Clearly it is necessary to introduce the sources

$$q_+(z,\mathbf{\Omega})=\mu\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger)\delta(z-\bar{z})-\mu\Psi_+(\bar{z}+\tau,\,\mathbf{\Omega};\,\bar{z},\,\mathbf{\Omega}^\dagger)\delta(z-(\bar{z}+\tau)),\quad (95)$$

$$q_-(z,\mathbf{\Omega})=-\mu\Psi_-(\bar{z},\mathbf{\Omega};\,\bar{z},\mathbf{\Omega}^\dagger)\delta(z-\bar{z}).$$

Applying the identity of Eq. (28) to $\Psi_\pm(z,\mathbf{\Omega};z_1,\mathbf{\Omega}_1)$ and $\tilde{\Psi}_\pm(z,\mathbf{\Omega};z_2,\mathbf{\Omega}_2)$ where $z_1+\tau\geq z_2\geq z_1$ results in

$$\mu_2\Psi_-(z_2,\mathbf{\Omega}_2;z_1,\mathbf{\Omega}_1)=\int d\mathbf{\Omega}\mu\tilde{\Psi}_-(z_2,\mathbf{\Omega};z_2,\mathbf{\Omega}_2)\Psi_+(z_2,\mathbf{\Omega};z_1,\mathbf{\Omega}_1)-\int d\mathbf{\Omega}\mu\Psi_+(z_1+\tau,\mathbf{\Omega};z_1,\mathbf{\Omega}_1)\times\tilde{\Psi}_-(z_1+\tau,\,\mathbf{\Omega};z_2,\,\mathbf{\Omega}_2).\quad (96a)$$

Assuming c independent of z we know from the "Principle of Invariance" that

$$\Psi(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)=\Psi(z-\bar{z},\,\mathbf{\Omega};0,\,\mathbf{\Omega}^\dagger).$$

If

$$S(\mathbf{\Omega},\mathbf{\Omega}')=\mu\Psi_-(0,\mathbf{\Omega};0,\mathbf{\Omega}'),\quad \tilde{S}(\mathbf{\Omega},\mathbf{\Omega}')=\mu\tilde{\Psi}_-(0,\mathbf{\Omega};0,\mathbf{\Omega}'),$$
$$T(\mathbf{\Omega},\mathbf{\Omega}')=\mu\Psi_+(\tau,\mathbf{\Omega};0,\mathbf{\Omega}'),\quad \tilde{T}(\mathbf{\Omega},\mathbf{\Omega}')=\mu\tilde{\Psi}_+(\tau,\mathbf{\Omega};0,\mathbf{\Omega}'),$$

it is possible to rewrite Eq. (96a) in the form

$$\mu\Psi_-(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)=\int d\mathbf{\Omega}'\tilde{S}(\mathbf{\Omega}',\mathbf{\Omega})\Psi_+(z,\mathbf{\Omega}';0,\mathbf{\Omega}_1)-\int d\mathbf{\Omega}'T(\mathbf{\Omega}',\mathbf{\Omega})\tilde{\Psi}_-(\tau-z,\,\mathbf{\Omega}';0,\,\mathbf{\Omega}_1).\quad (96b)$$

Putting $z=0$ and remembering the boundary condition of Eq. (94) one finds

$$S(\mathbf{\Omega},\mathbf{\Omega}_1)=\tilde{S}(\mathbf{\Omega}_1,\mathbf{\Omega}).\quad (96d)$$

Therefore, if $f(\mathbf{\Omega},\mathbf{\Omega}')$ has the time reflection symmetry S is symmetrical. In any event it is permissible to rewrite Eq. (96b) as

$$\mu\Psi_-(z,\mathbf{\Omega};0,\mathbf{\Omega}_1)=\int d\mathbf{\Omega}'S(\mathbf{\Omega},\mathbf{\Omega}')\Psi_+(z,\mathbf{\Omega}';0,\mathbf{\Omega}_1)-\int d\mathbf{\Omega}'T(\mathbf{\Omega}',\mathbf{\Omega})\tilde{\Psi}_-(\tau-z,\,\mathbf{\Omega}';0,\,\mathbf{\Omega}).\quad (96c)$$

Differentiate this equation with respect to z. Use the transfer equations to eliminate derivatives and put $z=0$. Remembering the boundary conditions and the definitions of S and T one obtains

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu_1}\right)S(\mathbf{\Omega}_0,\mathbf{\Omega}_1)=c\int\int d\mathbf{\Omega}'d\mathbf{\Omega}''\{\ \},$$

where

$$\{\ \}=f(\mathbf{\Omega}',-\mathbf{\Omega}'')\left[\delta_2(-\mathbf{\Omega}_0\cdot\mathbf{\Omega}')+\frac{S(\mathbf{\Omega}_0,\mathbf{\Omega}')}{\mu'}\right]\times\left[\delta_2(-\mathbf{\Omega}''\cdot\mathbf{\Omega}_1)+\frac{S(\mathbf{\Omega}'',\mathbf{\Omega}_1)}{\mu''}\right]-f(-\mathbf{\Omega}',\,\mathbf{\Omega}'')\frac{\tilde{T}(\mathbf{\Omega}',\mathbf{\Omega}_0)}{\mu'}\frac{T(\mathbf{\Omega}'',\mathbf{\Omega}_1)}{\mu''}.\quad (97)$$

[The same convention is used as in Eq. (36).]

The remaining conditions to determine S and T are obtained by comparing a solution with a beam incident on the slab from the left to one in which the incidence is from the right. Let $\tilde{\Phi}(z,\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)$ be the solution of the homogeneous Eq. (27) for $\bar{z}<z<\bar{z}+\tau$ subject to the boundary conditions

$$\tilde{\Phi}_-(\tau+\bar{z},\,\mathbf{\Omega};\,\bar{z},\,\mathbf{\Omega}^\dagger)=\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger),$$
$$\tilde{\Phi}_+(\bar{z},\mathbf{\Omega};\bar{z},\mathbf{\Omega}^\dagger)=0.$$

Again consider a solution of Eq. (27) everywhere which vanishes except for $\bar{z}<z<\bar{z}+\tau$ where it coincides with

[11] See reference 1, Chap. VII.

$\tilde{\Phi}$. The sources that must be introduced are

$$q_+(z,\mathbf{\Omega}) = -\mu\tilde{\Phi}_+(\bar{z}+\tau,\, \mathbf{\Omega};\, \bar{z},\, \mathbf{\Omega}^\dagger)\delta(z-(\bar{z}+\tau)),$$
$$q_-(z,\mathbf{\Omega}) = \mu\delta_2(\mathbf{\Omega}\cdot\mathbf{\Omega}^\dagger)\delta(z-(\bar{z}+\tau)) - \mu\tilde{\Phi}_-(\bar{z},\mathbf{\Omega};\, \bar{z},\mathbf{\Omega}^\dagger)\delta(z-\bar{z}). \quad (98)$$

Let $\psi_\pm^{(1)} = \Psi_\pm(z,\mathbf{\Omega};\, z_1,\mathbf{\Omega}_1)$ and $\tilde{\psi}_\pm^{(2)} = \tilde{\Phi}_\pm(z,\mathbf{\Omega};\, z_2,\mathbf{\Omega}_2)$ in the identity of Eq. (28) (where $z_1 \le z_2 \le z_1+\tau$). The result is

$$\int d\mathbf{\Omega}\mu\Psi_+(z_2,\mathbf{\Omega};\, z_1,\mathbf{\Omega}_1)\tilde{\Phi}_-(z_2,\mathbf{\Omega};\, z_2,\mathbf{\Omega}_2)$$
$$= \int d\mathbf{\Omega}\mu\tilde{\Phi}_-(z_1+\tau,\, \mathbf{\Omega};\, z_2,\, \mathbf{\Omega}_2)\Psi_+(z_1+\tau,\, \mathbf{\Omega};\, z_1,\, \mathbf{\Omega}_1). \quad (99)$$

Suppose $\tilde{\Psi}_\pm(z,\mathbf{\Omega})$ give a solution of Eqs. (27) with constant c. Clearly $\tilde{\Psi}'_\pm(z,\mathbf{\Omega}) = \tilde{\Psi}_\mp(-z,\, \Omega)$ is a solution of the equations obtained from Eq. (27) by replacing $f(\mathbf{\Omega},\mathbf{\Omega}')$ by $f(-\mathbf{\Omega},\, -\mathbf{\Omega}')$. Let us *assume* $f(\mathbf{\Omega},\mathbf{\Omega}') = f(-\mathbf{\Omega},\, -\mathbf{\Omega}')$. (This corresponds to invariance under spatial reflections. It is satisfied for all important applications.[12]) From the uniqueness theorem it follows that for constant c

$$\tilde{\Phi}_\mp(z,\mathbf{\Omega};\, \bar{z},\mathbf{\Omega}^\dagger) = \tilde{\Psi}_\pm(\bar{z}+\tau-z,\, \mathbf{\Omega};\, 0,\, \mathbf{\Omega}^\dagger). \quad (100)$$

Inserting this in Eq. (99) and using the translational invariance one concludes that

$$\int d\Omega\tilde{T}(\mathbf{\Omega},\mathbf{\Omega}_2)\Psi_+(z,\mathbf{\Omega};\, 0,\mathbf{\Omega}_1) = \int d\mathbf{\Omega}T(\mathbf{\Omega},\mathbf{\Omega}_1)\tilde{\Psi}_+(z,\mathbf{\Omega};\, 0,\mathbf{\Omega}_2). \quad (101)$$

Putting $z=0$ gives

$$\tilde{T}(\mathbf{\Omega}_1,\mathbf{\Omega}_2) = T(\mathbf{\Omega}_2,\mathbf{\Omega}_1). \quad (102)$$

Differentiate Eq. (101) with respect to z. Eliminate derivatives by means of the transfer equation and then put $z=0$. One obtains

$$\frac{T(\mathbf{\Omega}_0,\mathbf{\Omega}_1)}{\mu_0} - \frac{\tilde{T}(\mathbf{\Omega}_1,\mathbf{\Omega}_0)}{\mu_1} = c\int\int d\mathbf{\Omega}d\mathbf{\Omega}'\{\ \}, \quad (103)$$

where

$$\{\ \} = f(\mathbf{\Omega}',\, -\mathbf{\Omega})\frac{T(\mathbf{\Omega},\mathbf{\Omega}_1)}{\mu}\left[\delta_2(-\mathbf{\Omega}'\cdot\mathbf{\Omega}_0) + \frac{\tilde{S}(\mathbf{\Omega}',\mathbf{\Omega}_0)}{\mu'}\right] - f(\mathbf{\Omega},\, -\mathbf{\Omega}')\frac{\tilde{T}(\mathbf{\Omega},\mathbf{\Omega}_0)}{\mu}\left[\delta_2(-\mathbf{\Omega}'\cdot\mathbf{\Omega}_1) + \frac{S(\mathbf{\Omega}',\mathbf{\Omega}_1)}{\mu'}\right].$$

Finally, using the symmetry properties [Eqs. (96d) and (102)] we obtain as the fundamental equations for S and T

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu_1}\right)S(\mathbf{\Omega}_0,\mathbf{\Omega}_1) = c\int\int d\mathbf{\Omega}d\mathbf{\Omega}' f(\mathbf{\Omega},\, -\mathbf{\Omega}')\{\ \}_1, \quad (104a)$$

$$\left(\frac{1}{\mu_0}-\frac{1}{\mu_1}\right)T(\mathbf{\Omega}_0,\mathbf{\Omega}_1) = c\int\int d\mathbf{\Omega}d\mathbf{\Omega}'\{\ \}_2, \quad (105)$$

$$\{\ \}_1 = \left[\delta_2(-\mathbf{\Omega}_0\cdot\mathbf{\Omega}) + \frac{S(\mathbf{\Omega}_0,\mathbf{\Omega})}{\mu}\right] \times\left[\delta_2(-\mathbf{\Omega}'\cdot\mathbf{\Omega}_1) + \frac{S(\mathbf{\Omega}',\mathbf{\Omega}_1)}{\mu'}\right] - \frac{T(\mathbf{\Omega}_0,\mathbf{\Omega})}{\mu}\frac{T(\mathbf{\Omega}',\mathbf{\Omega}_1)}{\mu'},$$

and

$$\{\ \}_2 = f(\mathbf{\Omega}',\, -\mathbf{\Omega})\frac{T(\mathbf{\Omega},\mathbf{\Omega}_1)}{\mu}\left[\delta_2(-\mathbf{\Omega}'\cdot\mathbf{\Omega}_0) + \frac{S(\mathbf{\Omega}_0,\mathbf{\Omega}')}{\mu'}\right] - f(\mathbf{\Omega},\, -\mathbf{\Omega}')\frac{T(\mathbf{\Omega}_0,\mathbf{\Omega})}{\mu}\left[\delta_2(-\mathbf{\Omega}'\cdot\mathbf{\Omega}_1) + \frac{S(\mathbf{\Omega}',\mathbf{\Omega}_1)}{\mu'}\right].$$

It is remarkable that Eqs. (104a) and (105) have no explicit dependence on τ. (Exactly the same equations hold for slabs of all thicknesses.) Clearly some additional condition must be added to make the solution of the problem for a given slab unique. However, Eq. (105) tells, in essence, what the condition is. Solving for T in terms of the integral on the right it is seen that $T(\mathbf{\Omega}_0,\mathbf{\Omega}_1)$ is determined only up to a term proportional to $\delta(\mu_0-\mu_1)$. This must describe the direct contribution of the incident beam. Thus the correct form of Eq. (105) including the condition necessary to make Eqs. (104a) and (105) determinate is

$$T(\mathbf{\Omega}_0,\mathbf{\Omega}_1) = \mu_1\delta_2(\mathbf{\Omega}_0\cdot\mathbf{\Omega}_1)e^{-z/\mu_1} + \frac{c\mu_0\mu_1}{\mu_1-\mu_0}\int\int d\mathbf{\Omega}d\mathbf{\Omega}'\{\ \}_2. \quad (104b)$$

Equations (104a, b) are then the fundamental equations which determine the reflection and transmission for a slab. To recapitulate: we have assumed $c=$ constant (translational invariance), $f(\mathbf{\Omega},\mathbf{\Omega}') = f(-\mathbf{\Omega},\, -\mathbf{\Omega}')$ (spatial inversion invariance), but we have not assumed $f(\mathbf{\Omega},\mathbf{\Omega}') = f(-\mathbf{\Omega}',\, -\mathbf{\Omega})$ (time reflection invariance). If we do make this additional assumption we have also proved that S and T are symmetric. The essential difference between this and Chandrasekhar's[1] derivation is that we start from Eqs. (96c) and (101)—which follow immediately from the reciprocity identity and translational invariance. There is never any need to consider solutions of slab problems with different thicknesses.

APPENDIX

Since reciprocity relations are important for much more complicated situations than those treated above (where analytic, or semianalytic, solutions can be

[12] This is why the functions Φ_1, Φ_2 in Sec. V are usually even.

found), it seems worthwhile to sketch the proof of such theorems for the general case described by Eq. (1).

Let $\psi_i^{(1)}(\mathbf{r},\mathbf{\Omega})$ and $\tilde{\psi}_i^{(2)}(\mathbf{r},\mathbf{\Omega})$ satisfy

$$\mathbf{\Omega}\cdot\nabla\psi_i^{(1)}(\mathbf{r},\mathbf{\Omega})+\sigma_i(\mathbf{r})\psi_i^{(1)}=\sum_j\int F_{ij}(\mathbf{\Omega},\mathbf{\Omega}',r)\times\sigma_j(\mathbf{r})\psi_j^{(1)}(\mathbf{r},\mathbf{\Omega}')d\mathbf{\Omega}'+q_i^{(1)}(\mathbf{r},\mathbf{\Omega}),\quad\text{(A1)}$$

and

$$-\mathbf{\Omega}\cdot\nabla\tilde{\psi}_i^{(2)}(\mathbf{r},-\mathbf{\Omega})+\sigma_i(\mathbf{r})\tilde{\psi}_i^{(2)}=\sum_j\int F_{ji}(\mathbf{\Omega}',\mathbf{\Omega},\mathbf{r})\times\sigma_i(\mathbf{r})\tilde{\psi}_j^{(2)}(\mathbf{r},-\mathbf{\Omega}')d\mathbf{\Omega}'+q_i^{(2)}(\mathbf{r},-\mathbf{\Omega}).\quad\text{(A2)}$$

Multiply Eq. (A1) by $\tilde{\psi}_i^{(2)}(\mathbf{r},-\mathbf{\Omega})$, Eq. (A2) by $\psi_i^{(1)}(\mathbf{r},\mathbf{\Omega})$. Subtract, sum over i and integrate over the volume V under consideration and all $\mathbf{\Omega}$. The result is

$$\int dS\int d\mathbf{\Omega}\mathbf{n}_0\cdot\mathbf{\Omega}\sum_i\psi_i^{(1)}(\mathbf{r},\mathbf{\Omega})\tilde{\psi}_i^{(2)}(\mathbf{r},-\mathbf{\Omega})=\int d\mathbf{r}\int d\mathbf{\Omega}\sum_i[\tilde{\psi}_i^{(2)}(\mathbf{r},-\mathbf{\Omega})q_i^{(1)}(\mathbf{r},\mathbf{\Omega})-\psi_i^{(1)}(\mathbf{r},\mathbf{\Omega})q_i^{(2)}(\mathbf{r},-\mathbf{\Omega})].\quad\text{(A3)}$$

Specializing to various point sources and incident distribution it is possible to obtain a large number of relations between solutions of different problems.

If we define a matrix $\mathfrak{F}(\mathbf{\Omega},\mathbf{\Omega}',\mathbf{r})$ by

$$\mathfrak{F}_{ij}(\mathbf{\Omega},\mathbf{\Omega}',\mathbf{r})=F_{ij}(\mathbf{\Omega},\mathbf{\Omega}',\mathbf{r})\phi_j(\mathbf{r})$$

we might say that reciprocity in a strict sense exists provided there is a nonsingular matrix $Q(\mathbf{\Omega})$ such that

$$Q(\mathbf{\Omega})\mathfrak{F}^\dagger(-\mathbf{\Omega}',-\mathbf{\Omega},\mathbf{r})Q^{-1}(\mathbf{\Omega}')=\mathfrak{F}(\mathbf{\Omega},\mathbf{\Omega}',\mathbf{r}),\quad\text{(A4)}$$

where $\mathfrak{F}^\dagger$ denotes the transpose of $\mathfrak{F}$.

Under these conditions,

$$Q\tilde{\psi}=\psi,\quad\text{(A5)}$$

and the identities implied in Eq. (A3) will relate various solutions of the same equation. (A particular example of this is the reciprocity relation proved by Chandrasekhar[13] for the case of Rayleigh scattering of partially polarized light.)

[13] Reference 1, Chap. VII, Sec. 52.

Comparative Study of Potential Energy Functions for Diatomic Molecules

YATENDRA PAL VARSHNI
Department of Physics, Allahabad University, Allahabad, India

INTRODUCTION

THE correct representation of potential energy of a set of atoms as a function of the interatomic distance is of fundamental importance in molecular structure. A good deal of information about the structure of a molecule is summarized in its potential energy (P.E.) curves. Potential energy minima determine the bond lengths. The second derivatives of the P.E. with respect to distance give the force constants. And these determine the vibrational and rotational levels of the molecule. Anharmonicity constants depend on higher derivatives of the P.E. curve.

For molecular structure, the problem assumes its simplest form in the case of diatomic molecules where the interatomic distance is the one independent variable.

It is possible to plot the P.E. curves by making use of the energy levels themselves. Oldenberg (O1),[1] Rydberg (R6, R7), and Klein (K1) have developed methods of doing this, and Klein's method has been used with success by Rosenbaum (R5) and by Almy and Beiler (A1) for the upper electronic states of LiH and KH, respectively.

This method is very laborious and further it cannot be applied with accuracy at low vibrational quantum numbers for which the power series expansion method of Crawford and Jorgenson (C5) has to be applied. This power series for the potential energy in terms of the displacement of the atoms from the equilibrium position does not converge rapidly for any range of displacements and involves even more computation than the Rydberg-Klein method. Rees (R1, R2) has given an analytical formulation of the Klein-Rydberg method which enables the P.E. curve to be evaluated with accuracy in the region of the minimum and has applied it to certain states of Br_2.

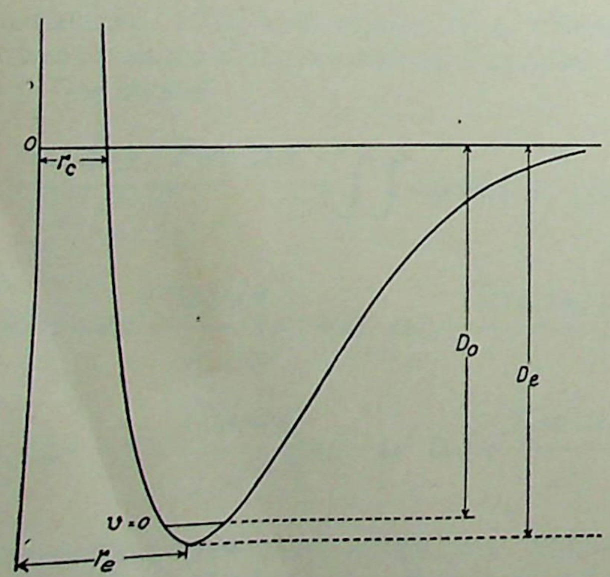


FIG. 1. Simple potential energy curve of a diatomic molecule.

On the theoretical side, by quantum-mechanical methods, it has not been found possible to calculate the P.E. curve to any great degree of accuracy due to mathematical and computational difficulties, except in the case of the simplest molecules like H_2^+ (Teller T3, Chakravarty C2), H_2 (James and Coolidge J1), etc. Classical electrostatic models for ionic molecules like alkali halides have yielded good results (Rittner R3, Varshni V2).

The most widely used method is to represent the P.E. curve by a suitable function—empirical or semi-empirical, and many functions have been suggested. First we formulate the criteria that a good potential function must satisfy. We are not concerned here with repulsive states. These criteria (Morse M6, Frost and Musulin F1, Lippincott and Schroeder L6) can be divided in two parts, (a) necessary and (b) desirable.

(a) Necessary:

(1) It should come asymptotically to a finite value as $r\rightarrow\infty$.

(2) It should have a minimum at $r=r_e$.

(3) It should become infinite at $r=0$. This need not be very strict, because results are practically the same if U becomes very large at $r=0$. Also if nuclear structure and forces are taken into account, this criterion is not exactly true. However, for our purposes, the change in the P.E. function at internuclear distances of 10^{-12} cm is of no consequence.

(b) Desirable:

(4) The above three criteria give a conventional form of P.E. curve as shown in Fig. 1. However, it is not the only possible form for the curve. Potential energy curves with at least one maximum between the main minimum and the dissociation limit are certainly known (Gaydon G1, Mulliken M7, M8, Herzberg H3, Barrow B3). Gaydon (G1) does not exclude the possibility of multiple maxima and minima.

The maximum often arises when an "attractive" potential curve in a low approximation is crossed by a "repulsive" potential curve. Because of the finer

[1] References are listed at the end of this article.

interaction, the intersection is avoided leading to a potential maximum of the lower of the resulting potential curves if the interaction is not too strong (Fig. 2). A few cases of such potential curves with maximum have been observed by means of predissociation phenomenon, for example for AlH and BH (Herzberg and Mundie H4). Maxima also arise due to Van der Waals interaction. Another case when the maximum is below the asymptote is shown in Fig. 3.

Thus to accommodate such states the potential function should be capable of giving rise to at least one maximum under certain conditions.

Frost and Musulin (F1) have given a theoretical discussion of the various criteria. In analogy with the wave-mechanical calculations, they consider the potential energy as the algebraic sum of two parts:— (i) the nuclear repulsive potential corresponding to merely a Coulomb potential $Z_1Z_2e^2/r$ where Z_1 and Z_2 are effective atomic numbers; (ii) the purely electronic energy defined as U_e which is also a function of r. Thus

$$U=\frac{Z_1Z_2e^2}{r}+U_e.$$

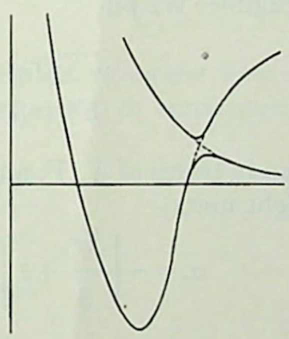


FIG. 2. Potential energy curve with one maximum. The maximum lies above the dissociation limit.

They have shown that an ideal potential function should also satisfy the following additional criteria. Somewhat similar ideas were advanced by Davidson and Price (D2) and Newing (N1) earlier.

(5) (a) U_e is finite at $r=0$.
(b) $U_e=U_e^0$ at $r=0$, where U_e^0 is the known "united" atom energy.

(6) $U_e\propto -Z_1Z_2e^2/r$ for large r.

(7) $\frac{dU_e}{dr}=0$ at $r=0$.

This condition is known to exist for H_2^+ as its electronic energy approaches the united atom He^+ (Matsen M3). This is believed to be correct also for other diatomic molecules, but the proof has apparently not been given.

(8) Van der Waals terms should introduce terms of the form $1/r^n$ (Coulson C4, Pauling and Beach P2).

Criteria 4 to 8 need not be exactly true.

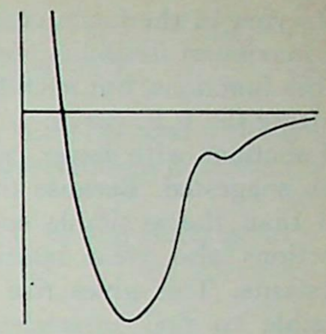


FIG. 3. Another case of potential energy curve with a maximum, the maximum lying below the dissociation limit.

The experimental constants which are available for fitting the curve are: 1. r_e—internuclear distance at equilibrium; 2. D_e—dissociation energy; 3. k_e—force constant; 4. α_e; 5. ω_ex_e; 6. ω_ey_e; 7. r_c—the critical distance, a quantity obtained indirectly from mass spectrometry appearance potentials. r_c is the value of r, less than r_e, at which $U=0$ or the same as at $r=\infty$; 8. U_e^0.

For most molecules only the first five are available. ω_ey_e is known only for a few of the more extensively investigated molecules, and that too not very accurately. So far, the r_c value is available only for H_2^+ (Stevenson S2). U_e^0 can only be used in conjunction with criterion 5b, which is a stringent condition and may not be satisfied by a function satisfactory from other criteria. Thus we are left with the first five constants.

Two types of analytical functions have been used: (1) power series of the type

$$U=c_1(r-r_e)+c_2(r-r_e)^2+c_3(r-r_e)^3\cdots, \qquad (1)$$

(2) closed formulas.

The first type has been investigated by Kratzer (K4) on old quantum theory and by Dunham (D5) and Chakravorti (C1) by wave mechanics. Dunham's work has been extended and simplified by Sandeman (S1).

Though this function has the advantage that the corresponding wave equation can be solved to a very good degree of accuracy, there is one serious objection. The series is not convergent and unsuitable for high r (the P.E. as obtained from this function tends to plus or minus infinity as r becomes infinite, instead of going to the dissociation limit), though it can be used to give a very accurate P.E. curve close to the minimum. Also this is not a very convenient expression from computational viewpoint. The only advantage that can be claimed for Dunham's expression is flexibility.

Numerous empirical or semiempirical functions have been proposed to represent the P.E. curve.

Comparative tests for some of the functions on a few molecules have been made (Lotmar L10, Coolidge, James, and Vernon C3, Davies D3). However, no extensive examination has been made so far to test the relative merits of the various functions.

In view of the fact that five constants are available, for maximum flexibility, one can keep five parameters in the functions, but such functions are only useful for plotting the P.E. curve.

Functions with fewer constants (3 or 4) have also been suggested. Because the number of parameters is less than the available spectroscopic constants, such functions also yield relations between the different constants. This gives rise to another question: is it possible to find interrelation between the different constants, irrespective of the type of binding, etc.? Special emphasis is laid on this point in this paper.

In principle we can utilize any of the known constants (in all equal to the number of parameters in the function) for fitting the function. However, this is not always satisfactory for two reasons: firstly, the expressions for the parameters in terms of constants usually are much simpler for r_e, k_e, D_e than for α_e and $\omega_e x_e$; secondly, the values of α_e and $\omega_e x_e$ are not known with the accuracy that the k_e and r_e values are known (at times values of α_e and/or $\omega_e x_e$ are not known at all). Hence, as a uniform policy we would determine the curve from k_e, r_e, and D_e when there are three constants. The function can be tested in two ways—(a) by comparing it with the experimental curve, (b) by evaluating the values of the coefficients in the series or the unused constants and comparing them with the actual values. Here we only utilize method (b).

Comparison of α_e and $\omega_e x_e$ is a more direct method, rather than of the coefficients of the series. Further it is more convenient for comparing different functions. Hence we adopt this method.

The condition that a satisfactory function should be capable of yielding the correct α_e and $\omega_e x_e$ values is only necessary but not sufficient. A function may yield correct α_e and $\omega_e x_e$, but still may not be satisfactory in the neighborhood of $r=0$. (See also Wu and Chao W2.)

To find the parameters from the experimental constants we have to solve the corresponding wave equation and compare the resulting coefficients of the terms with the series

$$G(v)=\omega_e(v+\tfrac{1}{2})-\omega_e x_e(v+\tfrac{1}{2})^2+\omega_e y_e(v+\tfrac{1}{2})^3+\cdots. \quad (2)$$

However, this is not always possible. The Schrödinger equation is exactly solvable in only a very limited number of cases (Manning M1, Eisenhart E1) and perturbation methods are often cumbersome. Hence the following procedure is adopted. The closed analytical function can be expanded as follows:

$$U(r)=\frac{1}{\lfloor 2}U''(r_e)(r-r_e)^2+\frac{1}{\lfloor 3}U'''(r_e)(r-r_e)^3+\frac{1}{\lfloor 4}U^{\text{iv}}(r_e)(r-r_e)^4+\cdots \quad (3)$$

which is the same as series (1) with

$$c_2=\frac{U''(r_e)}{\lfloor 2}$$

$$c_3=\frac{U'''(r_e)}{\lfloor 3}$$

$$c_4=\frac{U^{\text{iv}}(r_e)}{\lfloor 4}.$$

Dunham (D5) has shown that

$$\alpha_e=-\left[\frac{c_3}{c_2}r_e+1\right]\frac{6B_e^2}{\omega_e} \quad (4)$$

and

$$\mu\omega_e x_e=\left[\frac{15}{8}\left(\frac{c_3}{c_2}\right)^2-\tfrac{3}{2}\left(\frac{c_4}{c_2}\right)\right]\frac{h}{8\pi^2 c}, \quad (5)$$

where B_e = rotational constant and μ = reduced mass. Actually there are some further terms but they are negligible. We put

$$U'''(r_e)/U''(r_e)=X$$
$$U^{\text{iv}}(r_e)/U''(r_e)=Y.$$

Then in terms of X, Y, and μ_A (reduced mass in atomic weight units)

$$\alpha_e=-\left[\frac{Xr_e}{3}+1\right]\frac{6B_e^2}{\omega_e} \quad (6)$$

$$\omega_e x_e=\left[\frac{5}{3}X^2-Y\right]\frac{h}{64\pi^2 c\times 1.6597\times 10^{-24}\mu_A}$$

$$=\left[\frac{5}{3}X^2-Y\right]\frac{2.1078\times 10^{-16}}{\mu_A}$$

$$=\left[\frac{5}{3}X^2-Y\right]\frac{W}{\mu_A} \quad (7)$$

where $W=2.1078\times 10^{-16}$.

If we are dealing with a three constant curve we can fit the curve from k_e, r_e, and D_e and deduce the value of α_e and $\omega_e x_e$ from (6) and (7). The curve can be fitted from the following three conditions

$$U(r_e)-U(\infty)=-D_e \quad (8a)$$

$$\left(\frac{dU}{dr}\right)_{r=r_e}=0 \quad (8b)$$

and

$$\left(\frac{d^2U}{dr^2}\right)_{r=r_e}=k_e. \quad (8c)$$

The relative merits of the functions can be tested by calculating the values of α_e and $\omega_e x_e$ and comparing them with the experimental values. For a broad behavior of the function a new graphical method is used, which very clearly shows the applicability of a function without detailed calculations. In (6) and (7) it is only the quantities

$$\left[\frac{Xr_e}{3}+1\right] \text{ and } \left[\frac{5}{3}X^2-Y\right]$$

which will be different for different functions. For most of the functions

$$-\left[\frac{Xr_e}{3}+1\right]=F \tag{9}$$

$$\left[\frac{5}{3}X^2-Y\right]r_e^2=G \tag{10}$$

can be represented as functions of the dimensionless parameter Δ where

$$\frac{k_e r_e^2}{2D_e}=\Delta. \tag{11}$$

This Δ may be called *Sutherland parameter* in honor of Sutherland (S3) who was the first to emphasize its importance. Then

$$\alpha_e=F\cdot\frac{6B_e^2}{\omega_e} \tag{12}$$

$$\omega_e x_e=G\cdot\frac{W}{\mu_A r_e^2}. \tag{13}$$

From the later analysis, it appears that those functions which give α_e and $\omega_e x_e$ values higher than those given by the Morse function can be safely ruled out.

The following symbols are used:

$$\rho=r-r_e$$

$$\xi=\frac{r-r_e}{r_e}.$$

$A, B, C \cdots a, b, c \cdots m, n$, etc. are constants. Though often the constants in different functions are represented by the same symbols, the value of the constants for different functions is different and should not be confused.

DOUBLE-EXPONENTIAL AND MORSE FUNCTIONS

Assuming that the repulsive and attractive terms both vary in an exponential manner, the potential function can be put as

$$U=Ae^{-mr}-Be^{-nr}. \tag{14}$$

Morse (M6) suggested a special form of this:

$$U=D_e[1-e^{-a(r-r_e)}]^2. \tag{15}$$

This satisfies criteria (1) and (2) and although at $r=0$, U is not infinite, it is nevertheless a satisfactorily large quantity.

Applying conditions (8 a, b, c) one easily gets

$$a=(k_e/2D_e)^{\frac{1}{2}}.$$

Morse solved his function for a special case and found that

$$G(v)=\omega_e(v+\tfrac{1}{2})-\omega_e x_e(v+\tfrac{1}{2})^2$$

for the vibrational energy levels without any higher power of $(v+\frac{1}{2})$. However Ter Haar (T4) has shown that this conclusion is not rigorously true.

Pekeris (P4, P5) solved the Schrödinger equation for the Morse curve by a perturbation method and found that to a good approximation

$$\alpha_e=6B_e x_e\left[\left(\frac{B_e}{\omega_e x_e}\right)^{\frac{1}{2}}-\frac{B_e}{\omega_e x_e}\right]. \tag{16}$$

Since

$$U''(r_e)=2a^2D_e$$

$$U'''(r_e)=-6a^3D_e$$

$$U^{\text{iv}}(r_e)=14a^4D_e,$$

from Eqs. (6) and (7) we easily get

$$\alpha_e=(\Delta^{\frac{1}{2}}-1)\frac{6B_e^2}{\omega_e} \tag{17}$$

$$\omega_e x_e=8a^2\frac{W}{\mu_A}=8\Delta\frac{W}{r_e^2\mu_A}. \tag{18}$$

Equation (17) is equivalent to (16) by virtue of (18). Equation (16) has been tested by Pekeris and found to yield good results.

On the other hand (18) has been examined by Huggins (H5, H6), Linnett (L1), etc., and found to give very poor results.

Equation (18) can be simplified to

$$D_e=\frac{\omega_e^2}{4\omega_e x_e} \tag{19}$$

which corresponds to the linear Birge-Sponer extrapolation. Gaydon (G1) has shown that usually this method gives rather high values for the dissociation energies.

For hydrogen halides, Davies (D3) found the values of $U'''(r_e)$ and $U^{\text{iv}}(r_e)$ to be satisfactory.

This function predicts for H_2^+, $r_e=0.50$ A or 16% low as compared with Stevenson's (S2) experimental value of 0.595 A.

Badger (B2) has proposed the following empirical

rule for estimating "a" for single bonds:

$$a=1.15/(r_e-d_{ij})$$

where d_{ij} is the constant occurring in Badger's relation (B1) between k_e and r_e.

Huggins (H5, H6) modified the Morse function by putting a in (15) equal to $e^{mr_{12}}$, both r_{12} and m being the same for all electronic states of the same diatom. This gave satisfactory values of r_e from observed values of k_e and $\omega_e x_e$. But when he attempted to use the same function with the same values of r_{12} and m to calculate the heats of dissociation, the results were quite unsatisfactory, the errors being of the order of 100–200%. It appears that the function of Huggins gives an adequate representation of the relation between U and r in the neighborhood of r_e, but departs widely from the actual relation outside this region.

EXTENDED MORSE FUNCTION

Coolidge, James, and Vernon (C3) examined the following extended version of the Morse curve:

$$\frac{U}{D_e}=\sum_{n=2,3\cdots} c_n[1-e^{-a'(r-r_e)}]^n. \qquad (20)$$

In actual calculations on certain states of H_2, Coolidge, James, and Vernon have taken 7 terms in this series. The form is flexible, but it has one serious disadvantage: a' is difficult to determine, and in fact loses its significance as we take more terms in the extended Morse curve.

HULBURT-HIRSCHFELDER FUNCTION

Hulburt and Hirschfelder (H7) modified the simple Morse function in this way:

$$U=D_e[(1-e^{-x})^2+cx^3e^{-2x}(1+bx)] \qquad (21)$$

where

$$x=2\beta\xi=\frac{\omega_e}{2(B_eD_e)^{\frac{1}{2}}}\cdot\frac{r-r_e}{r_e}$$

and b and c are simple algebraic functions of the five spectroscopic constants. This has the great practical advantage that it uses just those five parameters which are most readily obtained from the study of band spectrum. Hulburt and Hirschfelder give a list of the numerical values of the parameters for 25 common diatomic molecules. With only a few exceptions, their potential curves lie above the Morse curves.

Recent work of Tawde and Gopalkrishnan (T1) on the transition probabilities in the C_2 (Swan) system suggests that this function is only satisfactory in the region $r \gtrsim r_e$.

RYDBERG FUNCTION

Rydberg (R6) proposed:

$$U=-D_e(1+b\rho)e^{-b\rho} \qquad (22)$$

which is similar to the radial wave function $R_{20}(r)$ of the hydrogen atom. This satisfies criteria (1) and (2), but like the Morse function, does not become ∞ at $r=0$, though it is sufficiently large at that point.

From the three conditions (8 a, b, c) one gets

$$b=(k_e/D_e)^{\frac{1}{2}}.$$

Further we have

$$U'''(r_e)=-2D_eb^3$$
$$U^{\text{iv}}(r_e)=3D_eb^4$$

whence we get

$$\alpha_e=\left[\frac{2\sqrt{2}}{3}\Delta^{\frac{1}{2}}-1\right]\frac{6B_e^2}{\omega_e} \qquad (23)$$

$$\omega_e x_e=\frac{22}{3}\Delta\cdot\frac{W}{r_e^2\mu_A}. \qquad (24)$$

In predicting transition probabilities of C_2 (Swan) system, the Rydberg function gives better performance than the Morse or Hulburt-Hirschfelder functions (Tawde and Gopalkrishnan T1).

ROSEN-MORSE FUNCTION

In its simple form the Rosen-Morse (R4) curve is

$$U=A\tanh(r/d)-C\,\text{sech}^2(r/d). \qquad (25)$$

This does not satisfy the 3rd criterion. The relations to observed constants are

$$\tanh(r_e/d)=-A/2C$$
$$D_e=(A+2C)^2/4C.$$

Let

$$f=1/C^2\cdot(4C^2-A^2)^2$$

then

$$k_e=U''(r_e)=f/8d^2C$$
$$U'''(r_e)=3Af/8d^3C^2$$
$$U(r_e)=(9A^2-8C^2)f/8d^4C^3.$$

We define another parameter p given by

$$r_e/d=p.$$

Then we have

$$\Delta=p^2(1+\tanh p)^2 \qquad (26)$$

$$\alpha_e=(2p\tanh p-1)\cdot\frac{6B_e^2}{\omega_e} \qquad (27)$$

$$\omega_e x_e=8p^2(3\tanh^2 p+1)\frac{W}{r_e^2\mu_A}$$

$$=8\Delta\frac{(3\tanh^2 p+1)}{(1+\tanh p)^2}\cdot\frac{W}{r_e^2\mu_A}$$

$$=8\Delta(1-e^{-2p}+e^{-4p})\frac{W}{r_e^2\mu_A}. \qquad (28)$$

Variation of F and G with p and Δ is tabulated in Table I.

Other methods of fitting the curve have been given by Lotmar (L10), Kronig (K6), and Davies (D3). Davies found that for hydrogen halides it gives results no better than Morse function. An extended four constant version of the Rosen-Morse function has been suggested by Lotmar (L10) which has been tested by Tawde and Gejji (T2).

MANNING-ROSEN-NEWING FUNCTION

The Manning-Rosen (M2) function is

$$U=\frac{1}{kd^2}\left[\frac{\beta(\beta-1)e^{-2r/d}}{(1-e^{-r/d})^2}-\frac{Ae^{-r/d}}{1-e^{-r/d}}\right] \tag{29}$$

which may be put as

$$U=-\frac{Be^{-r/d}+Ce^{-2r/d}}{(1-e^{-r/d})^2}. \tag{30}$$

It obeys the 3 necessary criteria. However, Manning and Rosen found that for this function "the dissociation energy . . . is greater than the quantity $\omega_e^2/4\omega_e x_e$ given by the Morse function." Thus

$$D_e=\frac{\omega_e^2}{4\omega_e x_e}+\delta$$

or

$$\omega_e x_e=\frac{\omega_e^2}{4(D_e-\delta)}>\frac{\omega_e^2}{4D_e}.$$

Thus it gives $\omega_e x_e$ values even higher than does the Morse function (which itself gives much higher values), and hence this function is not suitable.

Newing (N1, N2) studied the function

$$U=-D_e+D_e\left[\frac{(1-e^{-a\rho})e^{ar_e}}{e^{ar_e}-e^{-a\rho}}\right]^2. \tag{31}$$

(There is a misprint in the original equation as given in his paper. The first term in the denominator has been printed as e^{-ar_e}, whereas it should be e^{ar_e}.)

This expression on simplification takes the form of function (30).

PÖSCHL-TELLER FUNCTION

The function due to Pöschl and Teller (P6) has the form

$$U=A\ \mathrm{cosech}^2\alpha(r-r_0)-B\ \mathrm{sech}^2\alpha(r-r_0) \tag{32a}$$

which uses 4 constants.

We examine here the simpler version

$$U=M\ \mathrm{cosech}^2(r/d)-N\ \mathrm{sech}^2(r/d). \tag{32b}$$

This obeys the three criteria.

TABLE I. Rosen-Morse function.

p	Δ	F	G
0.7	1.26	−0.1539	8.208
0.8	1.77	0.0624	11.89
0.9	2.39	0.289	16.45
1.0	3.10	0.523	21.92
1.1	3.94	0.761	28.43
1.2	4.84	1.001	35.55
1.3	5.86	1.240	43.63
1.4	6.96	1.479	52.49
1.6	9.45	1.949	72.68
1.8	12.28	2.408	95.63
2.0	15.43	2.856	121.2
2.2	18.89	3.293	149.3
2.4	22.67	3.723	179.8
2.6	26.74	4.144	212.8
2.8	31.13	4.559	248.1

From conditions (8 a, b, c) we get the following relations (Davies D3):

$$\tanh^4(r_e/d)=M/N=y^4\quad\text{(say)}.$$

Then

$$d^2=8D_e/k_e\quad\text{or}\quad\Delta=4r_e^2/d^2$$

$$y=\tanh(r_e/d)$$

$$N=D_e/(1-y^2)^2.$$

Further

$$U''(r_e)=8N/d^2\cdot(1-y^2)^2$$

$$U'''(r_e)=-24N/yd^3\cdot(1-y^2)(1-y^4)$$

$$U^{\mathrm{iv}}(r_e)=\frac{8}{d^4}\left[15(MN)^{\frac{1}{2}}\left\{\frac{1-y^4}{y^2}\right\}^2-32N(1-y^2)^2\right]$$

whence we get

$$\alpha_e=[\Delta^{\frac{1}{2}}\coth\Delta^{\frac{1}{2}}-1]\frac{6B_e^2}{\omega_e} \tag{33}$$

$$\omega_e x_e=8\Delta\cdot\frac{W}{r_e^2\mu_A}. \tag{34}$$

While the value of α_e is different from the Morse function, the value of $\omega_e x_e$ comes out to be the same as in the Morse function.

Davies (D3) examined this function on hydrogen halides and from the values of $U'''(r_e)$ and $U^{\mathrm{iv}}(r_e)$ concluded that it is slightly superior to the Morse function.

HYLLERAAS FUNCTION

Hylleraas (H8, H9, H10) introduced the potential:

$$U=D_e\left[1-\frac{(1+a)(1+c)(z+b)}{(z+a)(z+c)(1+b)}\right] \tag{35}$$

where

$$z=e^{2(1+k)\beta\xi}$$

and the intermediate quantities a, b, c are defined by

$$a=(k-k_2)/(1+k_2)$$
$$b=\{k-(k_1+k_2)\}/(1+k_1+k_2)$$
$$c=(k-k_1)/(1+k_1)$$

k, k_1, k_2 being the parameters.

The function involves 6 parameters which are difficult to determine. Coolidge, James, and Vernon (C3) have given complicated equations for finding them. Obviously this function cannot furnish any relation between the molecular constants. This passes into the Morse curve upon making $k=k_1=0$, so that $a=b$, $c=0$. The Pöschl-Teller function is also a special case $k=0$, $k_1=\frac{1}{2}[\tanh(r/d)-1]$, $k_2=\frac{1}{2}[\coth(r/d)-1]$.

KRATZER FUNCTION

The Kratzer (K5) function approximates to the form

$$U=D_e\left[\frac{r-r_e}{r}\right]^2. \tag{36}$$

This has the drawback that it makes use of only two constants. The solution of the corresponding wave equation was first given by Fues (F3).

While it satisfies the three necessary criteria, application of conditions (8 a, b, c) leads to the abnormal result

$$\Delta=k_e r_e^2/2D_e=1. \tag{37}$$

This value is not found in any of the 23 molecules considered in the present paper. Even H_2 (which has the lowest value of Δ) has a value higher than 2, and it is extremely unlikely that any other molecule will satisfy this condition. Hence this function is not applicable to molecular problems. A modified form is considered later.

DAVIDSON FUNCTION

Davidson (D1) examined the function

$$U=a\left[\frac{r}{r_e}-\frac{r_e}{r}\right]^2. \tag{38}$$

This does not satisfy the criterion that U should be finite when $r=\infty$. Also only two constants are used. However, on applying conditions (8 a, b, c), this leads to the interesting result

$$k_e r_e^2=8a=\text{constant}. \tag{39}$$

Although this is not a universal relation, it is well satisfied by ground states of hydrogen halides (Kratzer K4, Glasstone G3, Davies D3, Heath, Linnett, and Wheatley H1).

It is easily deduced that

$$\alpha_e=2\cdot\frac{6B_e^2}{\omega_e} \tag{40}$$

$$\omega_e x_e=0. \tag{41}$$

F always comes out to be equal to 2, contrary to fact, and (40) does not hold for any molecule.

MECKE-SUTHERLAND FUNCTION

Mecke (M4) and Sutherland (S3) suggested

$$U=\frac{a}{r^m}-\frac{b}{r^n}. \tag{42}$$

Here both the attractive and the repulsive terms are of the reciprocal type. This type of function was used before by Grüneisen (G6), Mie (M5), etc., for forces in the solid state. A special case ($m=2$, $n=1$) has been treated by Fues (F3). Baughan (B4) used such a function for interaction between nonbonded atoms.

From conditions (8 a, b, c),

$$D_e=\frac{a}{r_e^m}\left(\frac{m-n}{n}\right)=\frac{b}{r_e^n}\left(\frac{m-n}{m}\right) \tag{43}$$

$$k_e=am(m-n)/r_e^{m+2}=bn(m-n)/r_e^{n+2} \tag{44}$$

whence

$$2\Delta=k_e r_e^2/D_e=mn. \tag{45}$$

Sutherland found that for the same class of molecules $k_e r_e^2/D_e$ is approximately constant, which led him to conclude that m and n are constant for similar molecules.

We can further deduce

$$\alpha_e=(m+n)\cdot\frac{2B_e^2}{\omega_e} \tag{46}$$

$$\omega_e x_e=\left[\frac{2}{3}m^2+\frac{7}{3}mn+\frac{2}{3}n^2+4(m+n)+4\right]\frac{W}{r_e^2\mu_A}. \tag{47}$$

From considerations of the relation between k_e and r_e, Linnett (L1) thinks that this function is not satisfactory.

Equation (42) is a four-constant function and we can determine their values from k_e, r_e, D_e, and α_e by (44), (45), and (46) and then try to evaluate $\omega_e x_e$ from (47) (actually we need calculate only m and n for finding $\omega_e x_e$). However, as we are restricting ourselves to three constant potentials we will not attempt it.

Sutherland (S4) later modified it by replacing r by $(r-d)$, so that it becomes

$$U=\frac{a}{(r-d)^m}-\frac{b}{(r-d)^n}. \tag{48}$$

This modified Sutherland function can be used (V3) to derive a series of very fruitful relations connecting the various molecular constants, by making certain assumptions regarding a, d, m, and n. These will be published later on.

Lovera (L11) applied (48) to 4b–6b group of molecules.

LINNETT FUNCTION

Linnett (L1, L2) investigated the function

$$U=\frac{a}{r^m}-be^{-nr}. \tag{49}$$

Conditions (8 a, b; c) provide the following relationships:

$$D_e=\frac{a}{r_e^m}\left[\frac{m-nr_e}{nr_e}\right] \tag{50}$$

$$k_e=am(m+1-nr_e)/r_e^{m+2} \tag{51}$$

$$k_er_e^{m+2}=ma+\frac{m^2D_er_e^m}{1+D_er_e^m/a}. \tag{52}$$

Assuming $m(=3)$ to be constant for the KK period of molecules, and a to be constant for all states of a given molecule, he showed that (52) gives a satisfactory relation between k_e and r_e. He further deduced

$$\omega_ex_e=\left[\frac{5}{3}\cdot\left\{\frac{20-(nr_e)^2}{4-nr_e}\right\}^2-\left\{\frac{120-(nr_e)^3}{4-nr_e}\right\}\right]\frac{2.098}{r_e^2\mu_A} \tag{53}$$

where r_e is in angstroms.

The mean percentage error in ω_ex_e values over the states considered was found to be 16, compared to 46 by the Morse function—a considerable improvement.

Linnett also traced regularities in the value of a and later (L2) extended the function to other periods. This function is capable of giving rise to a low maximum between $r=r_e$ and $r=\infty$, and such a behavior of the function was found with $^1\Sigma_g^+$ and $^3\Pi_g$ states of N_2, however, experimental evidence does not favor such a maximum in these states.

α_e and ω_ex_e can be deduced in terms of parameter t $(=nr_e)$ which can be found from

$$\Delta=\frac{m(m+1-t)t}{2(m-t)}. \tag{54}$$

Then

$$\alpha_e=\left[\frac{(m+1)(m+3)-t^2-t}{m+1-t}\right]\frac{6B_e^2}{\omega_e} \tag{55}$$

$$\omega_ex_e=\left[\frac{5}{3}\left\{\frac{(m+1)(m+2)-t^2}{m+1-t}\right\}^2-\left\{\frac{(m+1)(m+2)(m+3)-t^3}{m+1-t}\right\}\right]\frac{W}{r_e^2\mu_A}. \tag{56}$$

TABLE II. Linnett function.

t	Δ	F	G
0	0	6	11.67
1	2.25	7.33	27.2
1.5	3.75	8.1	37.28
2.0	6.0	9.0	50.6
2.25	7.875	9.53	59.4
2.5	11.25	10.17	70.47
2.6	13.65	10.46	75.92
2.7	17.55	10.78	82.16
2.75	20.625	10.95	83.69
2.8	25.2	11.13	89.96
3.0	∞	12	108.7

The expression for ω_ex_e coincides with that of Linnett for $m=3$.

Table II gives the variation of F and G with t and Δ for the special case of $m=3$.

WU-YANG FUNCTION

Wu and Yang (W3) used a function similar to the Born-Mayer function used for crystal forces:

$$U=ae^{-mr}-b/r^n. \tag{57}$$

This differs from the Linnett function in that the attractive and repulsive terms have been exchanged.

This satisfies only the first and second criteria. At $r=0$, this gives $U=-\infty$! Conditions (8 a, b, c) yield

$$k_er_e^{n+1}=-bn(n+1)/r_e+bnm. \tag{58}$$

For each molecular period Wu and Yang plotted $k_er_e^{n+1}$ against $1/r_e$ with different values of n. They obtained straight lines for a certain value of n, whence they concluded that n, m and b are constant in each period. This function will be examined in detail for α_e and ω_ex_e in a later publication.

PUPPI FUNCTION

Puppi (P7) derived a reduced functional relation, his reduced distance variable being

$$y=(r-r_e)/r_eD_e^{\frac{1}{2}}.$$

This was obtained by assuming a Morse function and introducing an empirical relation, equivalent to $k_er_e^2$ $=$const. Puppi's reduced equation is not universal as $k_er_e^2$ remains constant only within small groups of molecules. Furthermore, his variable y is complicated in that it mixes distance and energy.

LIPPINCOTT FUNCTION

Recently Lippincott (L3, L4) has given the following function for which he also gives a semitheoretical justification:

$$U=D_e[1-\exp(-n\rho^2/2r)][1+af(r)]. \tag{59}$$

Here $f(r)$ is a function of the internuclear distance, chosen such that $f(r)\to\infty$ when $r=0$ and $f(r)=0$ when

$r=\infty$. Lippincott says that for many purposes the $f(r)$ term is unimportant, and neglects it to a first approximation. Thus

$$U=D_e[1-\exp(-n\rho^2/2r)]. \qquad (60)$$

At $r=0$, r_e, and ∞, it has the values D_e, 0, and D_e, respectively.

Then from conditions (8 b, c) we get $n=k_e r_e/D_e$. Further

$$\alpha_e=-\tfrac{1}{2}\cdot\frac{6B_e^2}{\omega_e} \qquad (61)$$

$$\omega_e x_e=(6\Delta+3)\cdot\frac{W}{r_e^2\mu_A} \qquad (62)$$

$$=(3nr_e+3)\frac{W}{r_e^2\mu_A}.$$

Calculations of $\omega_e x_e$ have been reported by Lippincott and Schroeder (L6). They have calculated it in terms of n, where n is empirically given by

$$n=n_0(I/I_0)^{\frac{1}{2}}{}_A(I/I_0)^{\frac{1}{2}}{}_B$$

where $(I/I_0)_A$ and $(I/I_0)_B$ are the ionization potentials of atoms A and B, respectively, relative to those of the corresponding atoms in the same row and first column of the periodic table. For the H atom I/I_0 has been assigned the value 0.88 rather than 1. For diatomic alkali metal and alkali hydrides, $n_0=4.21\times10^8$, for most of the others $n_0=6.32\times10^8$.

For α_e, Lippincott and Schroeder report that (60) gives $\alpha_e=0$. [As (61) shows, this appears to be in error.]

Hence later they have used (59) with

$$af(r)=-a(r_e/r)^6[1-\exp(-b^2nr^{11}/2r_e^{12})]^{\frac{1}{2}}$$
$$+a(r_e/r)^{12}[1-\exp(-b^2nr^{11}/2r_e^{12})]$$

and deduced the following expressions

$$D_e=\omega_e^2/2nr_eB_e$$

$$\alpha_e=ab\left(\frac{nr_e}{2}\right)^{\frac{1}{2}}\cdot\frac{6B_e^2}{\omega_e}$$

$$\omega_e x_e=1.5B_e[0.25+nr_e/4+ab(nr_e/2)^{\frac{1}{2}}$$
$$+(5a^2b^2-ab^2)nr_e/2].$$

TABLE III. Frost-Musulin function.

s	Δ	F	G
0	0	0	3
1	1.5	0.555	15.29
2	4	1.17	34.42
3	7.5	1.8	60.60
4	12	2.44	93.96
5	17.5	3.09	134.6
5.5	20.625	3.47	157.6

Utilizing the fact that ab and b are found to be nearly constant for most molecules, they computed α_e and $\omega_e x_e$ from the above expressions. The calculated values agree well with the experimental values for the molecules considered.

The simple equation (60) has also been applied to the problem of the hydrogen bond by Lippincott and Schroeder (L5).

FROST-MUSULIN FUNCTION

Another recent function is due to Frost and Musulin (F1). From semitheoretical reasoning they have constructed

$$U=e^{-ar}(1/r-b). \qquad (63)$$

This satisfies the three criteria. Its special feature is that the united atom energy can be utilized in determining the constants a and b. Frost and Musulin found quite satisfactory results with this function for H_2^+ and H_2.

α_e and $\omega_e x_e$ are best represented in terms of a parameter s, defined by

$$\Delta=s^2/2+s \quad \text{or} \quad s=-1+(1+2\Delta)^{\frac{1}{2}}.$$

Then

$$\alpha_e=\left[\frac{2s^2+3s}{3(s+2)}\right]\frac{6B_e^2}{\omega_e} \qquad (64)$$

$$\omega_e x_e=\left[\frac{11s^4+66s^3+156s^2+144s+36}{3(s+2)^2}\right]\frac{W}{r_e^2\mu_A}. \qquad (65)$$

A short table of the variation of functions F and G with s and Δ is given in Table III.

Frost and Musulin (F2) have also attempted to obtain a "universal" potential energy function in terms of reduced potential energy U/D_e and reduced internuclear distance $(r-r_{ij})/(r_e-r_{ij})$, where r_{ij} is a constant for a given molecule formed from atoms i and j. They interpret r_{ij} as a measure of inner shell radii and correlate with Badger's d_{ij}. However, the results obtained for $U^{\text{III}}(r_e)$ and $U^{\text{IV}}(r_e)$ are rather poor, the average percentage deviations being 13.2 and 42, respectively, for the 23 molecules examined.

IONIC P.E. FUNCTIONS

For ionic molecules, P.E. functions based on a classical model have been used. These usually consist of several terms, including a Coulomb attractive term, a repulsive term and polarization terms etc. For details reference may be made to papers by Rittner (R1) and Varshni (V2).

SOME NEW POTENTIAL ENERGY FUNCTIONS

In the above, we have considered the various potential energy functions suggested by different authors.

Now we consider a few new potential energy functions and their possible applicability to diatomic molecules.

FIRST

$$U=D_e\{1-\exp[-b(r^2-r_e^2)]\}^2. \quad (66)$$

This satisfies criteria 1 and 2 and is sufficiently large at $r=0$.

We easily get

$$b=(k_e/8D_er_e^2)^{\frac{1}{2}}=\Delta^{\frac{1}{2}}/2r_e^2$$

$$\alpha_e=(\Delta^{\frac{1}{2}}-2)\frac{6B_e^2}{\omega_e} \quad (67)$$

$$\omega_ex_e=[8\Delta-12\Delta^{\frac{1}{2}}+12]\frac{W}{r_e^2\mu_A}. \quad (68)$$

Table IV shows the variation of F and G.

TABLE IV. First function.

Δ	F	G
0	−2	12
1	−1	8
2	−0.586	11.03
3	−0.268	15.22
4	0	20.00
6	+0.4495	30.61
8	0.8284	42.06
10	1.162	54.06
14	1.742	78.89
18	2.243	105.1
22	2.690	131.7
24	2.899	145.2
26	3.099	158.9

SECOND

$$U=D_e\left[1-\frac{r_e}{r}e^{-\alpha(r-r_e)}\right]^2. \quad (69)$$

This satisfies all the three criteria.

$$\alpha+\frac{1}{r_e}=\frac{\Delta^{\frac{1}{2}}}{r_e}$$

$$\alpha_e=\left[\Delta^{\frac{1}{2}}+\frac{1}{\Delta^{\frac{1}{2}}}-1\right]\frac{6B_e^2}{\omega_e} \quad (70)$$

$$\omega_ex_e=\left[8\Delta+12-\frac{8}{\Delta^{\frac{1}{2}}}+\frac{12}{\Delta}\right]\frac{W}{r_e^2\mu_A}. \quad (71)$$

Obviously α_e values by this function would be higher than those by Morse and because

$$\frac{12}{\Delta^{\frac{1}{2}}}\left[\Delta^{\frac{1}{2}}+\frac{1}{\Delta^{\frac{1}{2}}}\right]>\frac{8}{\Delta^{\frac{1}{2}}}.$$

TABLE V. Third function.

Δ	F	G
1	1.0	24
2	0.828	23.04
3	0.887	29.46
4	1.0	36.75
6	1.266	51.46
8	1.535	65.94
10	1.796	79.74
14	2.276	108.4
18	2.714	138.0
22	3.116	165.4
24	3.307	179.6
26	3.693	193.9

ω_ex_e values calculated by this function will also be higher than those by Morse.

THIRD

As the first function gave too low values, and second function high, hence a mixture was attempted

$$U=D_e\left\{1-\frac{r_e}{r}\exp[-\beta(r^2-r_e^2)]\right\}^2. \quad (72)$$

This also obeys the three criteria.

$$\beta=\frac{1}{2r_e^2}[\Delta^{\frac{1}{2}}-1]$$

$$\alpha_e=[\Delta^{\frac{1}{2}}+2/\Delta^{\frac{1}{2}}-2]\frac{6B_e^2}{\omega_e} \quad (73)$$

$$\omega_ex_e=[8\Delta-12\Delta^{\frac{1}{2}}+66-111/\Delta^{\frac{1}{2}}+73/\Delta]\frac{W}{r_e^2\mu_A}. \quad (74)$$

Table V depicts the variation of F and G with Δ.

FOURTH

$$U=B(A+e^{b/r})^2 \quad (75)$$

fulfills the three criteria.

Conditions (8 a, b, c) furnish

$$A=e^{b/r_e}$$

$$D_e=B(e^{b/r_e}-1)^2$$

$$b=r_e\ln A$$

$$\Delta=\left[\frac{\ln A}{1-1/A}\right]^2.$$

We get

$$\alpha_e=(\ln A+1)\frac{6B_e^2}{\omega_e} \quad (76)$$

$$\omega_ex_e=[8(\ln A)^2+24\ln A+64]\frac{W}{r_e^2\mu_A}. \quad (77)$$

As Δ is ~ 5 or higher, $(\ln A)^2\approx\Delta$. To this approxi-

TABLE VI. Sixth function.

Δ	F	G
1	−1	16
2	−0.293	15.66
3	+0.1546	20.62
4	0.5	27.00
6	1.041	41.27
8	1.474	56.33
10	1.846	71.77
14	2.475	103.0
18	3.007	134.5
22	3.477	166.3
24	3.695	182.1
26	3.903	198.0

mation

$$\alpha_e \approx [\Delta^{\frac{1}{2}}+1]\frac{6B_e^2}{\omega_e}$$

$$\omega_e x_e \approx [8\Delta+24\Delta^{\frac{1}{2}}+64]\frac{W}{r_e^2\mu_A}.$$

FIFTH

$$U=D_e\left[1-\left(\frac{r_e}{r}\right)^n\right]^2. \quad (78)$$

This is a generalization of Kratzer function and a special case of the Mecke-Sutherland function.

$$n^2=\Delta$$

$$\alpha_e=\Delta^{\frac{1}{2}}\frac{6B_e^2}{\omega_e} \quad (79)$$

$$\omega_e x_e=[8\Delta+12\Delta^{\frac{1}{2}}+4]\frac{W}{r_e^2\mu_A}. \quad (80)$$

SIXTH

$$U=D_e\left[1-\frac{r}{r_e}e^{-a(r-r_e)}\right]^2[1+Kf(r)] \quad (81)$$

where $f(r)$ is a function such that at $r=0$, $f(r)=\infty$ and at $r=\infty$, $f(r)=0$. This general function satisfies the three criteria. Neglecting the $f(r)$ term, we get

$$U=D_e\left[1-\frac{r}{r_e}e^{-a(r-r_e)}\right]^2. \quad (82)$$

This gives $U=D_e$ at $r=0$.
We have

$$ar_e=\Delta^{\frac{1}{2}}+1$$

$$\alpha_e=[\Delta^{\frac{1}{2}}-1/\Delta^{\frac{1}{2}}-1]\frac{6B_e^2}{\omega_e} \quad (83)$$

$$\omega_e x_e=[\Delta-12+8/\Delta^{\frac{1}{2}}+12/\Delta]\frac{W}{r_e^2\mu_A}. \quad (84)$$

Notice the similarity in coefficients of the terms in bracket in $\omega_e x_e$ with similar terms of the second function.

F and G values are recorded in Table VI.

SEVENTH

$$U=-Ar^n e^{-ar}[1-Kf(r)] \quad (85a)$$

where, as before, $f(r)$ is a function such that at $r=0$, $f(r)=\infty$ and at $r=\infty$, $f(r)=0$. This function is very similar to the Lippincott function.

As before, neglecting the $f(r)$ term

$$U=-Ar^n e^{-ar}. \quad (85b)$$

The general function (85a) satisfies the three criteria. However, the simplified version (85b) gives $U=0$ at $r=0$.

The values of the constants in (85b) can be obtained from conditions (8 a, b, c):

$$a=n/r_e$$

$$A=D_e/r_e^n e^{-n}$$

$$n=2\Delta.$$

α_e and $\omega_e x_e$ are given by

$$\alpha_e=-\tfrac{1}{3}\frac{6B_e^2}{\omega_e} \quad (86)$$

$$\omega_e x_e=[6\Delta+\tfrac{2}{3}]\frac{W}{r_e^2\mu_A}. \quad (87)$$

Here the F and G values are very similar to those of the Lippincott function; F comes out to be negative and G is only slightly less than Lippincott's value.

COMPARISON

For testing any potential function, we require accurate data for the various molecular constants. A search of the available experimental data reveals that there are only 23 neutral molecules for which reliable values are available. Even for these the situation is not quite satisfactory for all of them. Dissociation energies of N_2, CO, and NO are still an open question. Two values for N_2, three for CO, and two for NO have been widely discussed. Reference is made to recent reviews by Gaydon and Penney (G2), Long (L7, L8, L9), Pauling and Sheehan (P3), Glockler (G5), Valatin (V1), Gaydon (G1), and Douglas (D4). We have used all the values considered probable in our calculations. Recent experiments of Hendrie (H2) and Brook and Kaplan (B7) lend a strong support to the higher value (9.756 ev) for N_2.

Quantum-mechanical calculations by Kopineck (K2, K3) also support the high value of N_2. However, Wolfsberg (W1) considers that some of the approximations used by Kopineck are not justified and his

calculations should not be used as an argument favoring the 9.8 ev value for $D_e(N_2)$.

The dissociation energy for HF is also not certain. We have used the value given by Herzberg (H3). For Li_2 also there is slight uncertainty in the dissociation energy; again Herzberg's value has been followed.

The data for the 23 diatomic molecules for the ground states are given in Tables VII (a) and (b), collected from Herzberg (H3). For O_2, a more recent value of $D_0(=5.115$ ev) recommended by Brix and Herzberg (B5, B6) has been used. D_e has been calculated from the usual relation

$$D_e = D_0 + \tfrac{1}{2}\omega_e.$$

and k_e from $k_e = 4\pi^2\mu c^2\omega_e^2 = 5.8883\times10^{-2}\mu_A\omega_e^2$ dyne/cm.

Dissociation energies favored by Gaydon are marked with an asterisk. Most of the D_e values are given to 4 significant figures and some even to 5. These values are not that accurate. Because the conversion factors led to these values, and to keep consistency with original values, sometimes extra figures were retained.

For SO it is not yet certain that it is the ground state.

Some authors (e.g., Hulburt and Hirschfelder H7) use the "corrected" dissociation energies of molecules at absolute zero by subtracting the rotational, vibrational, and translational heat contents. Others have applied corrections to get the value of r_e at absolute zero (e.g., Verwey and de Boer V5). Such corrections are difficult to determine and often uncertain, hence we have completely omitted them.

While for most of the diatoms under consideration, the experimental evidence suggests a conventional potential energy curve, perhaps such is not the case with ZnH, CdH, and HgH. Dissociation energies determined by Birge-Sponer linear extrapolation come too high than the actual values. Mulliken (M7, M8) suspects that there might be a maximum and a long distance minimum in the P.E. curves of these molecules due to the approach of two interacting states. Glockler and Sister Horst (G4) have constructed a P.E. curve for HgH taking into account the fact that at large r, the molecule can change from a structure bonded by exchange forces to a quasi-molecule held together by London dispersion forces which leads to a small maximum.

The experimental values of F and G are given by

$$F = -\left[\frac{X\cdot r_e}{3}+1\right] = \frac{\alpha_e\omega_e}{6B_e^2} \qquad (88)$$

$$G = \left[\frac{5}{3}X^2 - Y\right]r_e^2$$

$$= \frac{\omega_e x_e r_e^2 \times 64\pi^2 c\times 1.6597\times10^{-24}\mu_A}{h}. \qquad (89)$$

TABLE VII.

(a)

Molecule figures in	μ_A	ω_e cm^{-1}	α_e cm^{-1}	$\omega_e x_e$ cm^{-1}
H_2	0.5041	4395.2	2.993	117.99
ZnH	0.9928	1607.6	0.2500	55.14
CdH	0.9992	1430.7	0.218	46.3
HgH	1.0031	1387.1	0.312	83.01
CH	0.930	2861.6	0.534	64.3
OH	0.9484	3735.2	0.714	82.81
HF	0.9573	4138.5	0.7705	90.069
HCl	0.9799	2989.7	0.3019	52.05
HBr	0.9956	2649.7	0.226	45.21
HI	1.0002	2309.5	0.183	39.73
Li_2	3.509	351.43	0.00704	2.592
Na_2	11.498	159.23	0.00079	0.726
K_2	19.488	92.64	0.000219	0.354
N_2 (i)	7.0038	2359.6	0.0187	14.456
N_2 (ii)*	7.0038	2359.6	0.0187	14.456
P_2	15.492	780.43	0.00142	2.804
O_2	8.000	1580.4	0.01579	12.073
SO	10.665	1123.7	0.00562	6.116
Cl_2	17.489	564.9	0.0017	4.0
Br_2	39.958	323.2	0.000275	1.145
I_2	63.466	214.6	0.000117	0.6127
ICl	27.422	384.18	0.000536	1.465
CO (i)	6.8584	2170.2	0.01748	13.46
CO (ii)	6.8584	2170.2	0.01748	13.46
CO (iii)*	6.8584	2170.2	0.01748	13.46
NO (i)	7.4688	1904	0.0178	13.97
NO (ii)*	7.4688	1904	0.0178	13.97

(b)

Molecule figures in	k_e 10^5 dynes/cm	r_e 10^{-8} cm	B_e cm^{-1}	D_e 10^{-12} erg
H_2	5.734	0.7417	60.809	7.607
ZnH	1.511	1.594	6.6794	1.523
CdH	1.204	1.762	5.437	1.228
HgH	1.137	1.740	5.549	0.740
CH	4.482	1.12	14.457	5.843
OH	7.792	0.971	18.871	7.340
HF	9.655	0.9171	20.939	10.65
HCl	5.157	1.275	10.591	7.394
HBr	4.117	1.414	8.473	6.277
HI	3.142	1.604	6.551	5.125
Li_2	0.2552	2.672	0.6727	1.685
Na_2	0.1717	3.078	0.1547	1.185
K_2	0.0985	3.923	0.0562	0.8322
N_2 (i)	22.96	1.094	2.010	12.046
N_2 (ii)*	22.96	1.094	2.010	15.863
P_2	5.556	1.894	0.3033	8.137
O_2	11.76	1.207	1.4457	8.351
SO	7.93	1.493	0.7089	6.522
Cl_2	3.279	1.988	0.2438	4.021
Br_2	2.457	2.284	0.08091	3.190
I_2	1.721	2.667	0.03736	2.489
ICl	2.296	2.321	0.11416	3.486
CO (i)	19.02	1.128	1.9314	14.865
CO (ii)	19.02	1.128	1.9314	15.602
CO (iii)*	19.02	1.128	1.9314	18.01
NO (i)	15.944	1.151	1.7046	8.673
NO (ii)*	15.944	1.151	1.7046	10.57

We have seen that F and G depend on the parameter Δ [defined in (11)]. It is these F and G values which differ for different functions. F determines α_e and G determines $\omega_e x_e$. The experimental values of Δ, F, and G have been tabulated in Table VIII. Figures 4 and 6 show respectively the actual values of F and G against Δ.

F and G values as given by various functions have been compared in Tables IX and X. Their variation

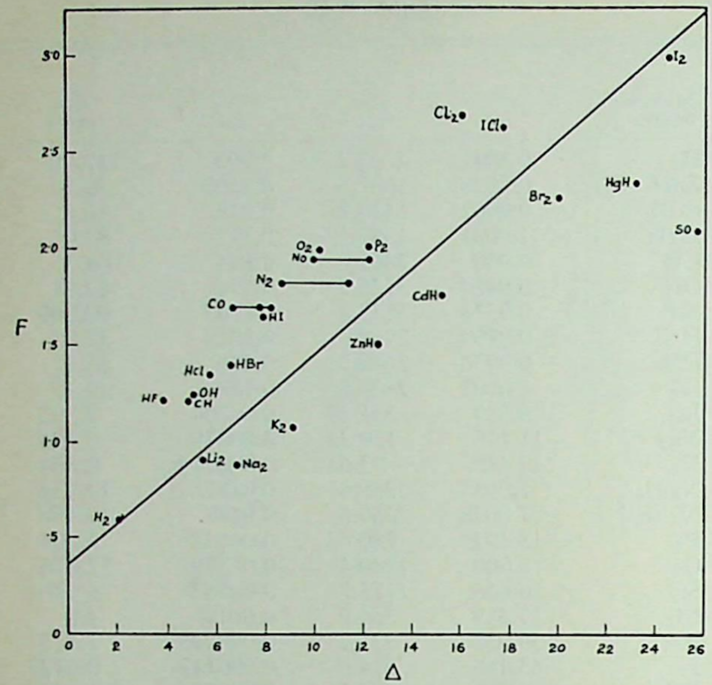


FIG. 4. Experimental values of the function F plotted against Sutherland parameter Δ.

with Δ for a few of the important functions have been shown graphically in Figs. 5 and 7, respectively. (To avoid confusion, the curves were not drawn in Figs. 4 and 6.)

Table VII (a) gives the experimental values of α_e and $\omega_e x_e$. Numerical calculations of α_e and $\omega_e x_e$ for a few of the functions along with percentage errors have been reported in Tables XI (a), (b) and XII (a), (b), (c), respectively. For calculating average percentage errors, only those values of N_2, CO, and NO which are favored

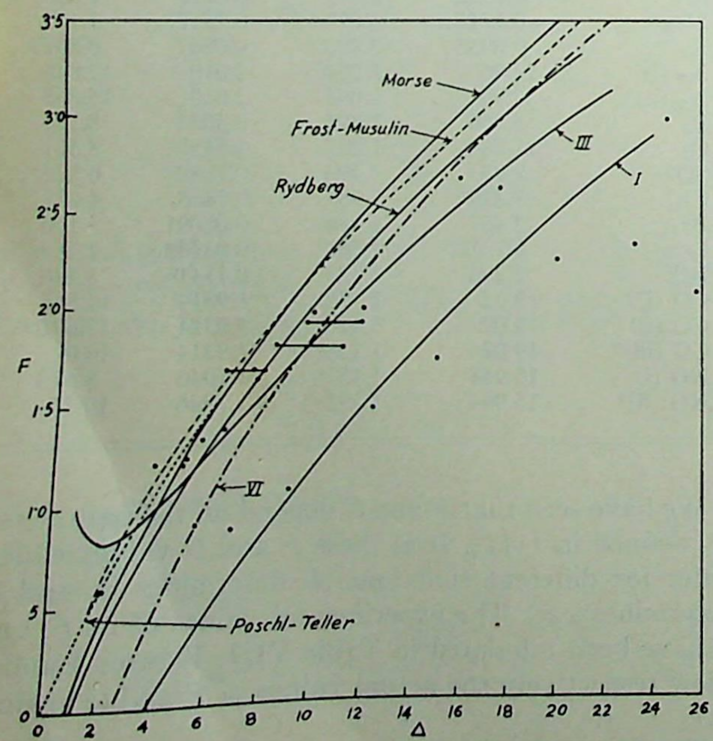


FIG. 5. Behavior of function F according to various potential energy functions. Experimental points are shown as dots.

by Gaydon (marked with an asterisk) have been taken into account.

DISCUSSION

A few important facts emerge clearly from Figs. 4 and 6. In case of both F and G, points are scattered. For F, they are rather too widely scattered. Had it been possible to represent all molecules by a "universal" P.E. curve, we should have obtained smooth curves passing through all the points. Thus the possibility that a "universal" P.E. curve exists can be ruled out. This agrees with what one expects theoretically. The shape of a P.E. curve depends on the electron configuration as well as on the position of neighboring electronic states and it is quite clear that any general relation cannot have exact validity. Of course, this is only relevant when

TABLE VIII.

Diatom	Δ	F	G
H_2	2.073	0.5928	15.52
ZnH	12.62	1.500	66.01
CdH	15.23	1.758	68.14
HgH	23.25	2.343	119.6
CH	4.814	1.218	35.58
OH	5.003	1.248	35.14
HF	3.813	1.213	34.40
HCl	5.669	1.342	39.33
HBr	6.554	1.390	42.69
HI	7.881	1.642	48.51
Li_2	5.408	0.9065	30.80
Na_2	6.864	0.8758	37.53
K_2	9.107	1.070	50.37
N_2 (i)	11.40	1.820	57.49
N_2 (ii)*	8.662	1.820	57.49
P_2	12.25	2.008	73.94
O_2	10.26	1.990	66.74
SO	25.81	2.093	131.5
Cl_2	16.12	2.693	131.2
Br_2	20.09	2.262	113.3
I_2	24.58	2.997	131.2
ICl	17.74	2.633	98.06
CO (i)	8.139	1.694	55.73
CO (ii)	7.752	1.694	55.73
CO (iii)*	6.719	1.694	55.73
NO (i)	12.18	1.944	65.58
NO (ii)*	9.995	1.944	65.58

we utilize only a limited number of molecular constants for fitting the P.E. curve, and try to find other constants from them. If the P.E. curve is made sufficiently flexible by including all the available molecular constants, no such question arises.

It may be thought that in view of the different nature of binding, HgH, CdH, and ZnH should be treated on somewhat different footing from other molecules for P.E. curve considerations. However, the position of these points in Figs. 4 and 6 does not seem to justify such an exclusion.

Nevertheless, Figs. 4 and 6 show that broadly speaking, F and G both increase with increasing Δ. This shows that it is possible to have "universal" potential curves broadly describing such a behavior, though not

TABLE IX.

	F values
Morse	$\Delta^{\frac{1}{2}}-1$
Rydberg	$2\sqrt{2}\Delta^{\frac{1}{2}}/3-1$
Rosen-Morse	$2p\tanh p-1$
Pöschl-Teller	$\Delta^{\frac{1}{2}}\coth\Delta^{\frac{1}{2}}-1$
Linnett	$(24-t^2)/(4-t)$
Lippincott	$-\frac{1}{2}$
Frost-Musulin	$(2s^2+3s)/3(s+2)$
I	$\Delta^{\frac{1}{2}}-2$
II	$\Delta^{\frac{1}{2}}+1/\Delta^{\frac{1}{2}}-1$
III	$\Delta^{\frac{1}{2}}+2/\Delta^{\frac{1}{2}}-2$
IV	$\Delta^{\frac{1}{2}}+1$
V	$\Delta^{\frac{1}{2}}$
VI	$\Delta^{\frac{1}{2}}-1/\Delta^{\frac{1}{2}}-1$
VII	$-\frac{1}{3}$

very accurately. It is with this point of view that we will consider the various functions.

We can also represent the variation of F and G with Δ empirically. As a first approximation, straight lines have been drawn in both cases. In case of G, the points are not so widely distributed and excepting a few points, a straight line would be a good approximation. However, in case of F one is tempted to draw two straight lines—one for the low-lying points and another for higher points. Then the question arises that for an unknown molecule, which line will be taken for prediction. On the lower line we may count H_2, Li_2, Na_2, K_2, ZnH, CdH, Br_2, HgH, SO. There does not seem to be any characteristic feature regarding such molecules for distinguishing them from others. While several of them belong to $s-s$ type of binding, and three of them are perhaps van der Waals molecules, there are still Br_2 and SO. Hence only one straight line was drawn. The

TABLE X.

	G values
Morse	8Δ
Rydberg	$22\Delta/3$
Rosen-Morse	$8\Delta(1-e^{-2p}+e^{-4p})$
Pöschl-Teller	8Δ
Linnett	$\frac{5}{3}\left[\frac{20-t^2}{4-t}\right]^2-\left[\frac{120-t^3}{4-t}\right]$
Lippincott	$6\Delta+3$
Frost-Musulin	$(11s^4+66s^3+156s^2+144s+36)/3(s+2)^2$
I	$8\Delta-12\Delta^{\frac{1}{2}}+12$
II	$8\Delta+12-8/\Delta^{\frac{1}{2}}+12/\Delta$
III	$8\Delta-12\Delta^{\frac{1}{2}}+66-111/\Delta^{\frac{1}{2}}+73/\Delta$
IV	$8\Delta+24\Delta^{\frac{1}{2}}+64$
V	$8\Delta+12\Delta^{\frac{1}{2}}+4$
VI	$8\Delta-12+8/\Delta^{\frac{1}{2}}+12/\Delta$
VII	$6\Delta+\frac{2}{3}$

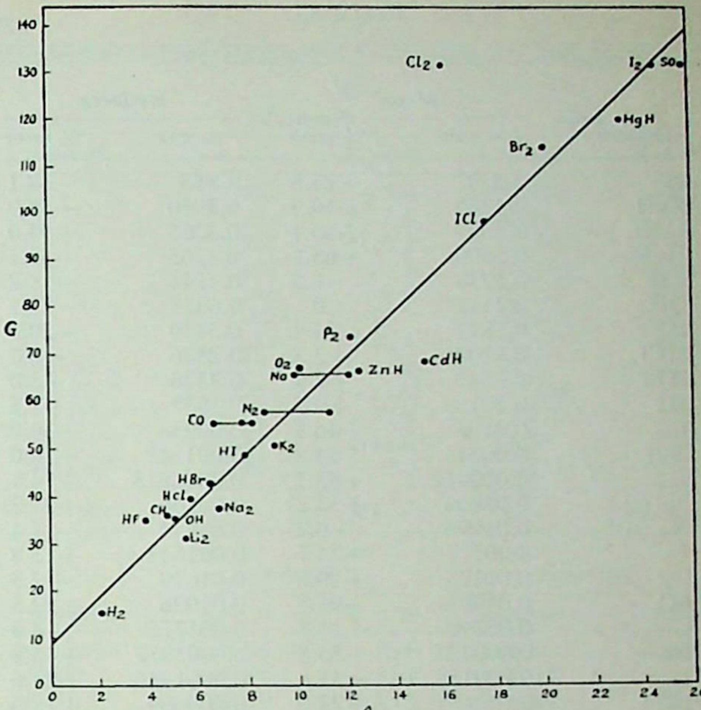


FIG. 6. Experimental values of the function G *versus* Sutherland parameter Δ.

equations of the two lines are:

$$F=0.11\Delta+0.36 \tag{90}$$

$$G=5\Delta+9. \tag{91}$$

One more question arises: While it is not possible to have exact "universal" P.E. function for *all* molecules, is it possible to have such a function for molecules with similar linkage, i.e., those belonging to the same

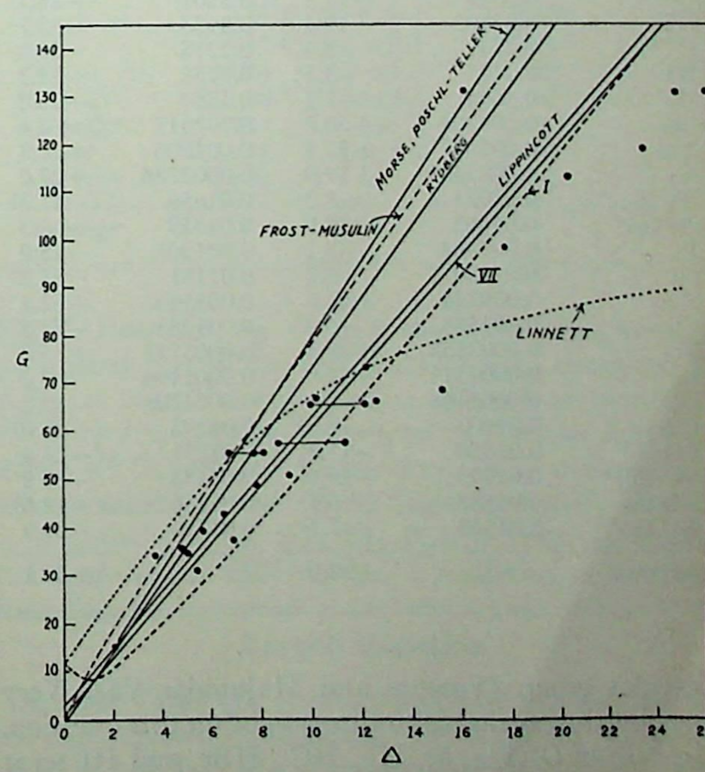


FIG. 7. Behavior of function G according to various potential energy functions. Experimental points are shown as dots.

TABLE XI.

(a)

Diatom	Morse α_e calc	Morse % error	Rydberg α_e calc	Rydberg % error
H_2	2.222	−25.8	1.853	−38.1
ZnH	0.4248	+69.9	0.3950	+58.0
CdH	0.3599	+65.1	0.3355	+54.0
HgH	0.5090	+63.1	0.4765	+52.7
CH	0.5238	−1.9	0.4744	−11.2
OH	0.7141	0	0.6425	−11.1
HF	0.5822	−24.4	0.5429	−29.5
HCl	0.3109	+3.3	0.2836	−6.0
HBr	0.2540	+7.9	0.2326	+3.0
HI	0.2015	+9.2	0.1857	+1.5
Li_2	0.0103	+46.3	0.00938	+30.0
Na_2	0.00146	+84.8	0.001342	+70.0
K_2	0.000412	+88.1	0.0003813	+74.1
N_2 (i)	0.02864	(+53.2)	0.02266	(+21.2)
N_2 (ii)*	0.01996	+6.2	0.01844	+1.4
P_2	0.00177	+24.7	0.001644	+15.8
O_2	0.00175	+10.8	0.01619	+2.5
SO	0.01095	+94.8	0.01026	+82.5
Cl_2	0.00190	+11.8	0.001775	+4.4
Br_2	0.000423	+53.8	0.0003957	+43.8
I_2	0.000154	+31.6	0.0001446	+23.6
ICl	0.000654	+22.0	0.000689	+13.6
CO (i)	0.01912	(+9.4)	0.01762	(+0.9)
CO (ii)	0.01844	(+5.5)	0.01694	(−3.0)
CO (iii)*	0.01643	−6.0	0.01506	−13.8
NO (i)	0.0228	(+28.1)	0.02117	(+18.1)
NO (ii)*	0.0198	+11.2	0.01832	+3.0
Average		±33.1		±28.0

(b)

Diatom	Third α_e calc	Third % error	Empirical α_e calc	Empirical % error
H_2	4.186	+39.8	2.969	−0.9
ZnH	0.3523	+40.9	0.291	+16.4
CdH	0.2996	+37.4	0.2524	+15.8
HgH	0.4311	+38.5	0.3886	+24.6
CH	0.4837	−9.4	0.3898	−27.0
OH	0.6319	−11.5	0.5207	−24.1
HF	0.6210	−19.4	0.4953	−35.7
HCl	0.2749	−8.9	0.2215	−26.6
HBr	0.2181	−3.5	0.1758	−22.2
HI	0.1694	−7.4	0.1369	−25.2
Li_2	0.009204	+30.7	0.007415	+4.1
Na_2	0.001247	+57.8	0.001006	+27.3
K_2	0.0003439	+57.0	0.0002786	+27.2
N_2 (i)	0.02023	(+8.2)	0.01659	(−11.3)
N_2 (ii)*	0.01665	−10.9	0.01349	−27.3
P_2	0.001464	+3.1	0.001207	−15.0
O_2	0.01450	−8.2	0.01181	−25.2
SO	0.009324	+65.8	0.008586	+52.8
Cl_2	0.001588	−6.6	0.001283	−24.5
Br_2	0.0003558	+29.1	0.0003127	+13.0
I_2	0.0001311	+12.0	0.0001196	+2.2
ICl	0.0005468	+2.1	0.0004706	−12.2
CO (i)	0.01603	(−8.3)	0.01293	(−26.0)
CO (ii)	0.01550	(−11.3)	0.01251	(−28.4)
CO (iii)*	0.01405	−19.6	0.01133	−35.2
NO (i)	0.01888	(+6.1)	0.01556	(−12.6)
NO (ii)*	0.01643	−7.7	0.01336	−24.9
Average		±22.9		±22.1

molecular group (Varshni and Majumdar V4)? Very few points are available to throw light on this question.

In case of G (Fig. 6), HF, HCl, HBr, and HI seem to be on a regular curve. The case with Li_2, Na_2, and K_2 is similar. Neglecting Cl_2 we can consider ICl, Br_2, and I_2 to be in a straight line. ZnH, CdH, and HgH are erratic which is not surprising in view of the irregular behavior of their molecular constants.

However, for F the situation is worse. Only for HF, HCl, HBr, and HI a regular curve can be drawn quite close to these points. Others do not show any regular behavior. In short, we can say that there is not sufficient evidence to give any definite conclusion on this point.

Now we consider the relative merits of each function.

Morse Function

This simple function though very widely used, to quote Partington (P1), "has been overworked in many branches of investigation," gives rather a poor performance.

α_e: (see Fig. 5, Table XI a): Excepting H_2, HF, and one point of CO, all other experimental points lie below the curve. Corresponding results are obtained in the numerical calculations which show the high average percentage error of 33.1.

$\omega_e x_e$: (see Fig. 7, Table XII a): The case here is similar. Only Cl_2, one point of CO, and HF lie above the Morse line. As expected, the average percentage error is high, being 31.2.

In a way, the Morse function sets an upper limit to F and G values. A satisfactory P.E. function should give F and G values which are lower than Morse's. As mentioned earlier, previous workers also found unsatisfactory results with this function. On the whole, this function is not suitable for using with a randomly chosen molecule.

Rydberg Function

From a comparison of F and G values of Morse and Rydberg functions, the latter gives lower values so the trend is in the right direction.

α_e: (see Fig. 5, Table XI a): There is a significant improvement from the Morse function, though even now only about 8 points lie above the curve. The average percentage error has slightly decreased to 28.

$\omega_e x_e$: (see Fig. 7, Table XII a): Though better than Morse, still only about 5 points are above the Rydberg line. The average percentage error has decreased to 23.1.

On the whole, this function is distinctly better than the Morse function.

Rosen-Morse Function

α_e: (see Table I): From the expression for F when p is small (i.e., Δ is small), this gives lower values than Morse function, as p increases (i.e., Δ increases), the Rosen-Morse F approaches the F of the Morse function.

$\omega_e x_e$: (see Table I): Behavior of G is very similar to that of F.

Thus, the Rosen-Morse function is somewhat better than Morse function, though more complicated.

TABLE XII.

(a)

Diatom	Morse, Poschl-Teller $\omega_e x_e$ calc	Morse, Poschl-Teller % error	Rydberg $\omega_e x_e$ calc	Rydberg % error
H_2	126.3	+7.0	115.8	−1.9
ZnH	84.35	+52.9	77.32	+40.2
CdH	82.80	+78.8	75.90	+63.9
HgH	129.2	+55.6	118.4	+42.6
CH	69.58	+8.2	63.78	−5.5
OH	94.36	+13.9	86.5	+4.4
HF	79.86	−11.3	73.2	−18.7
HCl	60.00	+15.3	55.00	+5.6
HBr	55.53	+22.8	50.9	+12.6
HI	51.64	+30.0	47.34	+24.2
Li_2	3.639	+37.3	3.336	+24.8
Na_2	1.048	+44.3	0.9605	+32.3
K_2	0.512	+44.6	0.4692	+32.5
N_2 (i)	22.93	(+58.0)	21.02	(+45.4)
N_2 (ii)*	17.42	+20.5	15.97	+10.4
P_2	3.717	+32.5	3.408	+21.5
O_2	14.85	+23.0	13.61	+12.7
SO	9.615	+57.2	8.814	+44.1
Cl_2	3.932	−1.7	3.604	−9.9
Br_2	1.624	+41.8	1.489	+30.0
I_2	0.9183	+49.8	0.8418	+37.4
ICl	2.101	+43.4	1.926	+31.5
CO (i)	15.76	(+17.1)	14.45	(+7.4)
CO (ii)	14.98	(+11.3)	13.73	(+2.0)
CO (iii)*	12.99	−3.5	11.91	−11.5
NO (i)	20.75	(+48.5)	19.02	(+36.1)
NO (ii)*	17.04	+21.9	15.62	+11.9
Average		±31.2		±23.1

(b)

Diatom	Lippincott $\omega_e x_e$ calc	Lippincott % error	First $\omega_e x_e$ calc	First % error
H_2	117.6	−0.3	86.14	−27.0
ZnH	65.77	+19.3	58.76	+6.5
CdH	64.13	+38.5	59.12	+27.7
HgH	98.93	+19.2	97.27	+17.2
CH	57.59	−10.4	47.31	−26.4
OH	77.84	−6.0	59.37	−28.3
HF	67.76	−24.7	49.94	−44.5
HCl	48.97	−5.9	36.77	−29.3
HBr	44.81	−0.9	35.70	−21.0
HI	41.19	+3.7	33.88	−14.7
Li_2	2.983	+15.1	2.218	−14.4
Na_2	0.8547	+17.7	0.6863	−5.5
K_2	0.4051	+14.4	0.3418	−3.4
N_2 (i)	17.96	(+24.2)	15.77	(+9.1)
N_2 (ii)*	13.82	−4.4	11.56	−20.0
P_2	2.902	+3.4	2.578	−8.1
O_2	11.68	−3.3	10.24	−15.2
SO	7.352	+20.2	7.335	+1.9
Cl_2	3.041	−23.4	2.83	−29.3
Br_2	1.248	+9.0	1.202	+5.0
I_2	0.7028	+14.7	0.6965	+13.7
ICl	1.621	+10.6	1.531	+4.5
CO (i)	12.52	(−7.0)	10.43	(−22.5)
CO (ii)	11.96	(−11.1)	9.806	(−27.1)
CO (iii)*	10.45	−22.4	8.370	−37.8
NO (i)	16.21	(+16.0)	14.39	(+3.0)
NO (ii)*	13.42	−4.0	11.51	−17.6
Average		±12.7		±18.2

(c)

Diatom	Seventh $\omega_e x_e$ calc	Seventh % error	Empirical $\omega_e x_e$ calc	Empirical % error
H_2	99.06	−16.0	147.5	+25.8
ZnH	63.81	+15.7	60.23	+9.2
CdH	62.38	+35.4	57.86	+24.9
HgH	97.27	+17.2	86.96	+4.7
CH	53.38	−16.8	59.74	−7.1
OH	72.33	−12.6	80.19	−3.2
HF	61.65	−31.5	73.48	−18.4
HCl	45.88	−11.8	49.42	−5.1
HBr	42.36	−6.3	44.23	−2.2
HI	39.27	−1.2	39.64	−0.2
Li_2	2.786	+7.4	3.033	+17.0
Na_2	0.8091	+11.5	0.8362	+15.2
K_2	0.3887	+9.8	0.3832	+8.2
N_2 (i)	17.37	(+20.0)	16.60	(+14.8)
N_2 (ii)*	13.23	−8.5	13.15	−9.0
P_2	2.813	+0.3	2.664	−5.0
O_2	11.26	−6.7	10.90	−9.7
SO	7.243	+18.4	6.428	+5.1
Cl_2	2.970	−25.7	2.732	−31.7
Br_2	1.226	+7.1	1.107	−3.3
I_2	0.6763	+10.4	0.616	+0.5
ICl	1.586	+8.3	1.446	−1.3
CO (i)	11.96	(−11.1)	12.00	(−10.9)
CO (ii)	11.40	(−15.3)	11.53	(−14.3)
CO (iii)*	9.9	−26.5	8.172	−39.3
NO (i)	15.71	(+12.5)	14.89	(+6.6)
NO (ii)*	12.93	−7.5	12.56	−10.1
Average		±13.6		±11.1

Pöschl-Teller Function

α_e: (see Fig. 5): Since $\coth\Delta^{\frac{1}{2}} > 1$, it will always give higher values than the Morse function. However, as Δ increases, this almost merges with the Morse value.

$\omega_e x_e$: (see Fig. 7, Table XII a): Same expression, as given by the Morse function.

Thus this function is slightly inferior to the Morse function, though the difference is not large. The fact that Davies (D3) found it slightly superior to the Morse function for HF, HCl, and HI, is because he compared $U'''(r_e)$ and $U^{\text{iv}}(r_e)$, rather than α_e and $\omega_e x_e$. $U'''(r_e)$ depends on α_e, while $U^{\text{iv}}(r_e)$ depends both on α_e and $\omega_e x_e$. As Tables VI and IX of Davies' paper show, he obtained $U'''(r_e)$ for the Pöschl-Teller higher than $U'''(r_e)$ for the Morse. This is in agreement with our Fig. 5. However, for $U^{\text{iv}}(r_e)$, the effect of α_e and $\omega_e x_e$ got mixed up. While the effect of $\omega_e x_e$ remained the same on both the functions, α_e effected an increase in $U^{\text{iv}}(r_e)$ for the Pöschl-Teller over the Morse. As Davies found $U^{\text{iv}}(r_e)$ for the Morse lower than experimental, he concluded that the Pöschl-Teller is slightly superior.

Actually only for HF, the Pöschl-Teller is superior to the Morse; for HCl and HI, it is inferior, as is evident from Fig. 5.

Linnett Function

α_e: (see Table II): F values given by this function are too high. Indeed the smallest value of $F=6$ (for $\Delta=0$) given by this function is almost twice the highest value of F found in the 23 molecules under consideration. Hence for α_e it is quite unsatisfactory.

$\omega_e x_e$: (see Fig. 7, Table II): The course of the G curve is very peculiar. For low Δ it is greater even than the Morse, for high Δ it is too low.

At least for $m=3$, this is not a very useful function.

Simple Lippincott Function

α_e: It gives a negative value which is meaningless.

$\omega_e x_e$: (see Fig. 7, Table XII b): From the expression for $\omega_e x_e$, for $\Delta > 2$ (which is usually the case) the values are much lower than from the Morse function (almost three-fourth of the Morse function). Table XII b shows that the average percentage error is much lower (± 12.7) as compared to the Morse (± 31.2) and the Rydberg (± 23.1) functions, indicating a significant improvement.

The simple function (60) does not satisfy the third criterion; to achieve this the $f(r)$ term has to be introduced. The $f(r)$ given by Lippincott and Schroeder makes the function a five constant function and thus rather complicated. Only by assuming that ab and b are constant can one predict α_e and $\omega_e x_e$. As we are confining ourselves to three-constant functions, calculations for α_e and $\omega_e x_e$ have not been carried out.

From the spread of points in Fig. 7, when two or more D_e have been suggested for a diatom, there is no reason to assume the value favored by this function as the "correct" one. Thus, Lippincott's statement that "where two or more alternative values have been proposed, this function should suggest which is the correct one" seems to be rather optimistic.

Frost-Musulin Function

α_e: (see Fig. 5, Table III): For low Δ it is slightly higher than Morse function, but for high Δ it is slightly lower than Morse function.

$\omega_e x_e$: (see Fig. 7, Table III): Behavior of G is very similar to that of F.

On the whole this function is very close to the Morse function, though slightly more complex.

First Function

α_e: (see Fig. 5, Table IV): This gives much lower values for F than the Morse function, which is also evident from Fig. 5.

$\omega_e x_e$: (see Fig. 7, Tables IV and XII b): Usually gives lower values. Average percentage error (18.2) is lesser than that for the Morse and Rydberg functions.

Second Function

As shown before, both for α_e and $\omega_e x_e$ it would give higher values than the Morse function and hence is unsuitable.

Third Function

α_e: (see Fi. 5, Tables V and XI b): The average percentage error (22.9) is significantly lower than that by Morse (3.1) and Rydberg (28.0) functions.

$\omega_e x_e$: (see Table V): At low Δ (but >2), it is midway between the Linnett and Frost-Musulin curves; as Δ increases it almost follows the Frost-Musulin curve (to avoid overlapping, the curve has not been drawn).

Fourth Function

This gives much higher values for F and G than the Morse function, and hence can be dismissed.

Fifth Function

This also gives higher values than the Morse for both F and G and hence is not applicable.

Sixth Function

α_e: (see Fig. 5, Table VI): Course of the F curve is encouraging. However, for $\Delta < 2.5$, it is negative.

$\omega_e x_e$: (see Table VI): Very close to the Rydberg function.

Seventh Function

α_e: It gives a negative value of F $(=-\frac{1}{3})$ which is absurd.

$\omega_e x_e$: (see Fig. 7, Table XII c): Slightly lower than, and parallel to the Lippincott function. Consequently, the average percentage error (13.6) is also very near to that of Lippincott (12.7).

The results given by this function have a striking resemblance to those of the Lippincott function.

Empirical Relations

α_e: (Fig. 4, Table XI b): As can be expected, the average percentage error is lowest, being 22.1. However, this is not far from the third function (22.9).

$\omega_e x_e$: (Fig. 6, Table XII c): Here also the average percentage error is lowest, being 11.1.

CONCLUSIONS

We may summarize as follows:

It is not possible to find three-constant "universal" potential energy functions, and consequently, there are no "universal" relations connecting α_e and $\omega_e x_e$ with Δ.

Several functions give relations which may be used for estimating α_e and $\omega_e x_e$. Further, $\omega_e x_e$ can be estimated to a greater degree of accuracy than α_e. A surprising fact emerges that a function which may be useful for calculating $\omega_e x_e$, may fail completely to reproduce α_e (e.g. Lippincott, Seventh).

On the basis of these 23 diatoms, for α_e, the functions in order of increasing accuracy are:

(i) Morse, Pöschl-Teller, Rosen-Morse, Frost-Musulin
(ii) Rydberg
(iii) Third, Sixth

And for $\omega_e x_e$:
(i) Morse, Pöschl-Teller, Rosen-Morse, Frost-Musulin, Third, Sixth
(ii) Rydberg
(iii) Lippincott, Seventh.

Thus for over-all representation of the P.E. curves, the Rydberg, Third, and Sixth functions are the most useful.

Such diverse looking functions as those of Morse, Pöschl-Teller, Rosen-Morse, and Frost-Musulin are actually extremely close to each other.

The best values of α_e and $\omega_e x_e$ are obtained from

$$\alpha_e = (0.11\Delta + 0.36)6B_e^2/\omega_e \tag{92}$$

$$\omega_e x_e = (5\Delta + 9)W/r_e^2\mu_A. \tag{93}$$

Equation (93) can be transformed as

$$D_e = \frac{5Wk_e r_e^2}{2\omega_e x_e r_e^2 \mu_A - 9W}. \tag{94}$$

For estimating dissociation energies, (94) deserves to be employed in preference to $D_e = \omega_e^2/4\omega_e x_e$. However, this cannot serve as a strong argument for favoring any of the values when several appear to be possible (e.g. NO, CO, N_2, etc.).

In Table VIII the Sutherland parameter Δ varies regularly in a molecular group (V4) and increases with increasing μ_A.

The author is thankful to Professor K. Banerjee for his kind interest in the work and the Government of India for a research grant.

REFERENCES

A1. G. M. Almy and A. C. Beiler, Phys. Rev. **61**, 476 (1942).
B1. R. M. Badger, J. Chem. Phys. **2**, 128 (1934).
B2. R. M. Badger, J. Chem. Phys. **3**, 710 (1935).
B3. R. F. Barrow, J. Chem. Phys. **22**, 573 (1954).
B4. E. C. Baughan, Quart. Revs. **7**, 103 (1953).
B5. P. Brix and G. Herzberg, J. Chem. Phys. **21**, 2240 (1953).
B6. P. Brix and G. Herzberg, Can. J. Phys. **32**, 110 (1954).
B7. M. Brook and J. Kaplan, Phys. Rev. **96**, 1540 (1954).
C1. S. K. Chakravorti, Z. Physik **109**, 25 (1938).
C2. S. K. Chakravarty, Phil. Mag. **28**, 423 (1939).
C3. Coolidge, James, and Vernon, Phys. Rev. **54**, 726 (1938).
C4. C. A. Coulson, Proc. Roy. Soc. (Edinburgh) **61**, 20 (1941).
C5. F. H. Crawford and T. Jorgenson, Phys. Rev. **49**, 745 (1936).
D1. P. M. Davidson, Proc. Roy. Soc. (London) **A135**, 459 (1932).
D2. P. M. Davidson and W. C. Price, Proc. Roy. Soc. (London) **130**, 105 (1930).
D3. M. Davies, J. Chem. Phys. **17**, 374 (1949).
D4. A. E. Douglas, J. Phys. Chem. **59**, 109 (1955).
D5. J. L. Dunham, Phys. Rev. **41**, 713, 721 (1932).
E1. L. P. Eisenhart, Phys. Rev. **74**, 87 (1948).
F1. A. A. Frost and B. Musulin, J. Chem. Phys. **22**, 1017 (1954).
F2. A. A. Frost and B. Musulin, J. Am. Chem. Soc. **76**, 2045 (1954).
F3. E. Fues, Ann. Physik **80**, 376 (1926).
G1. A. G. Gaydon, *Dissociation Energies and Spectra of Diatomic Molecules* (Chapman and Hall, London, 1953).
G2. A. G. Gaydon and W. G. Penney, Proc. Roy. Soc. (London) **A183**, 374 (1945).
G3. S. Glasstone, *Recent Advances in Physical Chemistry* (Churchills, 1938), p. 206.
G4. G. Glockler and Sister H. V. Horst, J. Chem. Phys. **20**, 1448 (1952).
G5. G. Glockler, Ann. Revs. Phys. Chem. **3**, 151 (1952).
G6. E. Grüneisen, Ann. Physik **26**, 393 (1908).
H1. Heath, Linnett, and Wheatley, Trans. Faraday Soc. **46**, 137 (1950).
H2. J. M. Hendrie, J. Chem. Phys. **22**, 1503 (1954).
H3. G. Herzberg, *Spectra of Diatomic Molecules* (D. Van Nostrand Company, Inc., Princeton, 1950).
H4. G. Herzberg and L. G. Mundie, J. Chem. Phys. **8**, 263 (1940).
H5. M. L. Huggins, J. Chem. Phys. **3**, 473 (1935).
H6. M. L. Huggins, J. Chem. Phys. **4**, 308 (1936).
H7. H. M. Hulburt and J. O. Hirschfelder, J. Chem. Phys. **9**, 61 (1941).
H8. E. A. Hylleraas, Z. Physik **96**, 643 (1935).
H9. E. A. Hylleraas, J. Chem. Phys. **3**, 595 (1935).
H10. E. A. Hylleraas, Physik. Z. **36**, 599 (1935).
J1. H. M. James and A. S. Coolidge, J. Chem. Phys. **1**, 825 (1933).
K1. O. Klein, Ann. Physik **76**, 226 (1932).
K2. H. J. Kopineck, *Conference on Quantum Mechanical Methods in Valence Theory* (Office of Naval Research, Washington, 1951), p. 45.
K3. H. J. Kopineck, Z. Naturforsch. **7a**, 314 (1952).
K4. A. Kratzer, Z. Physik **3**, 289 (1920).
K5. A. Kratzer, Ann. Physik **67**, 127 (1922).
K6. R. de L. Kronig, *The Optical Basis of the Theory of Valency* (Cambridge University Press, New York, 1935), p. 88.
L1. J. W. Linnett, Trans. Faraday Soc. **36**, 1123 (1940).
L2. J. W. Linnett, Trans. Faraday Soc. **38**, 1 (1942).
L3. E. R. Lippincott, J. Chem. Phys. **21**, 2070 (1953).
L4. E. R. Lippincott, J. Chem. Phys. **23**, 603 (1955).
L5. E. R. Lippincott and R. Schroeder, J. Chem. Phys. **23**, 1099 (1955).
L6. E. R. Lippincott and R. Schroeder, J. Chem. Phys. **23**, 1131 (1955).
L7. L. H. Long, Proc. Roy. Soc. (London) **A198**, 62 (1949).
L8. L. H. Long, Research **3**, 291 (1950).
L9. L. H. Long, Experientia **7**, 195 (1951).
L10. W. Lotmar, Z. Physik **93**, 528 (1935).
L11. G. Lovera, Nuovo cimento **8**, 1014 (1951).
M1. M. F. Manning, Phys. Rev. **48**, 161 (1935).
M2. M. F. Manning and N. Rosen, Phys. Rev. **44**, 953 (1933).
M3. F. A. Matsen, J. Chem. Phys. **21**, 928 (1953).
M4. R. Mecke, Z. Physik **42**, 390 (1927).
M5. G. Mie, Ann. Physik **11**, 657 (1903).
M6. P. M. Morse, Phys. Rev. **34**, 57 (1929).
M7. R. S. Mulliken, Phys. Rev. **51**, 310 (1937).
M8. R. S. Mulliken, J. Phys. Chem. **41**, 5 (1937).
N1. R. A. Newing, Phil. Mag. **19**, 759 (1935).
N2. R. A. Newing, Phil. Mag. **29**, 298 (1940).
O1. O. Oldenberg, Z. Physik **76**, 226 (1932).
P1. J. R. Partington, *An Advanced Treatise on Physical Chemistry*, (Longmans Green and Company, New York, 1949), Vol. I, p. 327.
P2. L. Pauling and J. Y. Beach, Phys. Rev. **47**, 686 (1935).
P3. L. Pauling and W. F. Sheehan, Proc. Natl. Acad. Sci. (U. S.) **83**, 143 (1949).
P4. C. L. Pekeris, Phys. Rev. **44**, 953 (1933).
P5. C. L. Pekeris, Phys. Rev. **45**, 98 (1934).
P6. G. Pöschl and E. Teller, Z. Physik **83**, 143 (1933).
P7. G. Puppi, Nuovo cimento **3**, 338 (1946).

R1. A. L. G. Rees, Proc. Phys. Soc. (London) **49**, 998 (1947).
R2. A. L. G. Rees, Proc. Phys. Soc. (London) **49**, 1008 (1947).
R3. E. S. Rittner, J. Chem. Phys. **19**, 1030 (1951).
R4. N. Rosen and P. M. Morse, Phys. Rev. **42**, 210 (1932).
R5. E. J. Rosenbaum, J. Chem. Phys. **6**, 16 (1938).
R6. R. Rydberg, Z. Physik **73**, 376 (1931).
R7. R. Rydberg, Z. Physik **80**, 514 (1933).
S1. I. Sandeman, Proc. Roy. Soc. (Edinburgh) **60**, 210 (1940).
S2. D. P. Stevenson, Vortex **12**, 198 (1951).
S3. G. B. B. M. Sutherland, Proc. Indian Acad. Sci. **8**, 341 (1938).
S4. G. B. B. M. Sutherland, J. Chem. Phys. **8**, 161 (1940).
T1. N. R. Tawde and K. Gopalkrishnan, Indian J. Phys. **28**, 469 (1954).
T2. N. R. Tawde and N. V. Gejji, Indian J. Phys. **30**, 299 (1956).
T3. E. Teller, Z. Physik **61**, 458 (1930).
T4. D. Ter Haar, Phys. Rev. **70**, 222 (1946).
V1. J. G. Valatin, Proc. Phys. Soc. (London) **A66**, 1185 (1953).
V2. Y. P. Varshni, Trans. Faraday Soc. **53**, 132 (1957).
V3. Y. P. Varshni (unpublished).
V4. Y. P. Varshni and K. Majumdar, Indian J. Phys. **29**, 38 (1955).
V5. E. J. W. Verwey and J. H. de Boer, Rec. trav. chim. **59**, 633 (1940).
W1. M. Wolfsberg, J. Chem. Phys. **21**, 2166 (1953).
W2. C. K. Wu and S. C. Chao, Phys. Rev. **71**, 118 (1947).
W3. C. K. Wu and C. T. Yang, J. Phys. Chem. **48**, 295 (1944).

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 4 OCTOBER, 1957

Energy Levels of Light Nuclei ($Z=11$ to $Z=20$). II.

P. M. ENDT, *Fysisch Laboratorium der Rijksuniversiteit, Utrecht, Netherlands*

AND

C. M. BRAAMS, *Stichting voor Fundamenteel Onderzoek der Materie, Utrecht, Netherlands*

INTRODUCTION

THE present compilation is a modernized version of the paper on the same subject[1] which appeared in 1954. It contains the experimental material published in the period from December 1, 1953, up to about February 1, 1957, and together with En 54a provides a fairly complete bibliography for the years prior to 1957. In a few cases, reference was made to papers which appeared after February 1, 1957, either because the authors had sent preprints, or because the importance of the paper made it desirable to add a note in the proof.

Presentation of some subjects which were treated in more detail in En 54a has been shortened. Little attention is given to cross-section measurements. For neutron cross sections the reader is referred to the compilation by Hughes and Harvey.[2] Many-particle reactions, as $(n,2n)$ or spallation reactions, are omitted altogether. Theoretical papers are quoted only briefly, and the bibliography is certainly incomplete on theoretical subjects.

The situation regarding nuclear masses has much improved since the appearance of the 1954 review because of the accurate mass-spectroscopic measurement of the S^{32} mass by Quisenberry, Scolman, and Nier,[3] and of the recent extension of Q-value chains into the $A>32$ region,[4] connecting nuclides up to $A=45$ to S^{32} through precision Q-value measurements. In Table LXV the masses are listed which were used to compute the Q_m values quoted after the heading of each specific reaction. The procedure followed in arriving at the values listed, is outlined in the text preceding the table.

Rules adopted for construction of the level diagrams and the systematics of arrangement of the material in the text are not significantly different from those in En 54a.

(a) Levels known with an accuracy of 15 kev or better, and found preferably from two or more reactions, are marked by heavy lines, doubtful levels by broken lines. Three-decimal figures are reserved for levels known with an accuracy of 5 kev or better.

(b) Resonance energies indicated in the level diagrams are given in the laboratory system.

(c) Nuclides are presented in order of increasing A, and nuclides of the same A in order of increasing Z.

(d) Generally each nuclear reaction is treated under the heading of the *final* nucleus, exceptions are reactions where resonances are observed, which are treated under the compound nucleus, and the β decay of unstable nuclei, treated under the parent nucleus, and then as the first reaction. The order of the other reactions is determined by the initial nucleus starting with the lowest-Z element and within that element with the lowest-A isotope.

ACKNOWLEDGMENT

We are particularly grateful to the many physicists who sent results before publication.

Na^{20}

(not illustrated)

I. $Na^{20}(\beta^+)Ne^{20}$ $Q_m=15.3$*

Positron decay proceeds at least partly to states of Ne^{20} between 6.8 and 10.8 Mev which decay by α emission. Half-life, 0.25 sec, as measured by α detection (Al 50a), and 0.23 ± 0.08 sec as measured by β^+ detection (Sh 51b). Another value, 0.385 ± 0.01 sec, is reported (Aj 52). Energy of α's is >2 Mev (Al 50a).

Theoretical prediction of Na^{20} spin, De 53a.

Discussion of possibility that positron decay is superallowed, Bo 55.

II. $Ne^{20}(p,n)Na^{20}$ $Q_m=-16.1$

Threshold 16.9 Mev, Al 50a.

Na^{21}

(Fig. 1)

I. $Na^{21}(\beta^+)Ne^{21}$ $Q_m=3.52$

Half-life, 23 ± 2 sec (Cr 40) and 22.8 ± 0.5 sec (Sc 52); also Bo 53. The β^+ end point is 2.50 ± 0.03 Mev determined with a 180° spectrometer (Sc 52); also Bo 53. A γ ray reported (Po 40), but recent work shows no γ ray of $E_\gamma>0.5$ Mev (Sc 52). Log $ft=3.6$, confirming that spin of Na^{21} is same as that of mirror nucleus Ne^{21}, namely $3/2^+$.

---

[1] P. M. Endt and J. C. Kluyver, Revs. Modern Phys. **26**, 95 (1954) (quoted as En 54a).

[2] D. J. Hughes and J. A. Harvey, "Neutron cross sections," Brookhaven National Laboratory BNL 325 (1955).

[3] Quisenberry, Scolman, and Nier, Phys. Rev. **102**, 1071 (1956).

[4] Endt, Buechner, Braams, Paris, and Sperduto, Phys. Rev. **106**, 764 (1957).

* For each reaction the Q value calculated from the masses is designated Q_m.

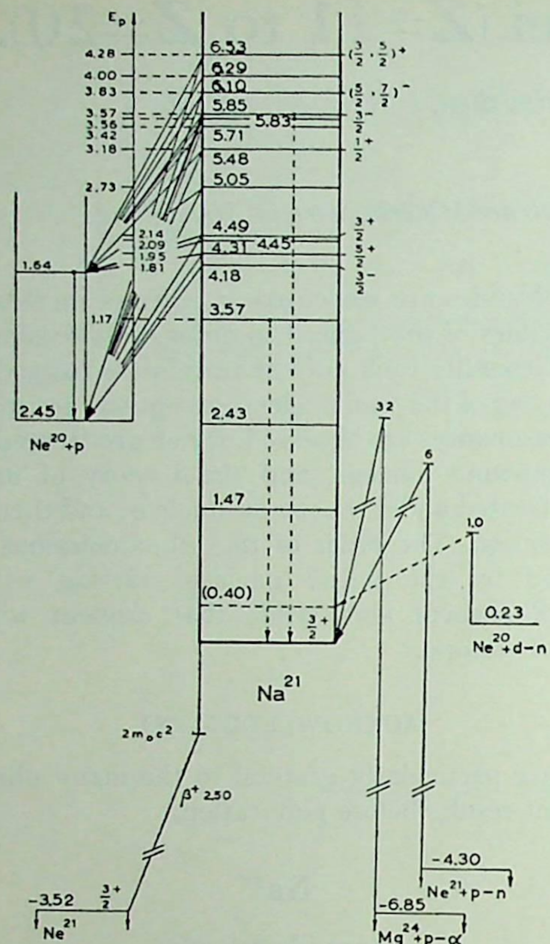


FIG. 1.

II. $Ne^{20}(p,\gamma)Na^{21}$ $Q_m = 2.45$

Weak resonance at $E_p = 1.165$ Mev found in γ-ray yield for protons in 0.5–1.3-Mev region bombarding a thin isotopic Ne^{20} target (Br 47). This corresponds to level in Na^{21} at 3.57 Mev. For protons in 1.35—4.4-Mev region, resonances for production of annihilation radiation from Na^{21} decay found at $E_p = 2.15$, (3.45), 3.56, and 3.57 Mev (Co 54b, Va 53).

III. $Ne^{20}(p,p)Ne^{20}$ $E_b = 2.45$
$Ne^{20}(p,p'\gamma)Ne^{20}$ $E_b = 2.45$ $E_x = 1.63$

Yield of elastically scattered protons measured at six angles from 90° to 167.5°. Energies, widths, spins, and parities as determined by phase-shift analysis in Table I (Ha 53, Ha 55a).

Resonances for inelastic proton scattering leading to 1.63 Mev level of Ne^{20} generally correspond to those for elastic scattering; resonances at $E_p = 2.09$ and 4.00 Mev, found from inelastic scattering only. Cross sections and angular distributions of inelastically scattered protons and 1.63-Mev γ radiation (Ga 53, Co 54b).

IV. $Ne^{20}(d,n)Na^{21}$ $Q_m = 0.23$

Slow-neutron thresholds indicate Q values of -1.24 ± 0.02 and -2.201 ± 0.007 Mev, corresponding to levels at 1.47 and 2.43 Mev (Ma 56b). At $E_d = 1.0$ Mev, recoil protons in nuclear emulsions indicate $Q = -0.17 \pm 0.05$ Mev (Sw 52), corresponding to a level at 0.40 Mev.

TABLE I. Resonances in $Ne^{20} + p$

E_p (Lab) (Mev)	E_x (Mev)	decay [e]	Γ (kev)	J, π
1.165[a]	3.57	γ[a]		
1.81[d]	4.18	p_0[d]	180[d]	3/2−[d]
1.953[d]	4.31	p_0[d] p_1[b]	6[d]	5/2+[d]
2.09[b]	4.45	p_1[b]		
2.135[d]	4.49	γ[c] p_0[d] p_1[b]	17[d]	3/2+[d]
2.73[b]	5.05	p_0[d] p_1[b]	double[d]	
3.176[d]	5.48	p_0[d] p_1[b]	110[d]	1/2+[d]
3.42[d]	5.71	p_0[d] p_1[b]	≈20[d]	
3.552[d]	5.83	γ[c] p_0[d] p_1[b]	≈2[d]	
3.566[d]	5.85	γ[c] p_0[d] p_1[b]	25[d]	3/2−[d]
3.828[d]	6.10	p_0[d] p_1[b]	6[d]	(5/2, 7/2)−[d]
4.00[b]	6.29	p_1[b]		
4.28[d]	6.53	p_0[d] p_1[b]	150[d]	(3/2, 5/2)+[d]

[a] Br 47. [b] Co 54b. [c] Va 53. [d] Ha 55a. [e] Resonances in yield of capture γ rays or in yield of Na^{21} activity indicated by γ, resonances for elastic scattering, by p_0, and resonances in yield of 1.63 Mev γ ray from first excited state of Ne^{20}, by p_1.

V. $Ne^{21}(p,n)Na^{21}$ $Q_m = -4.30$

Observed, En 54a.

VI. $Mg^{24}(p,\alpha)Na^{21}$ $Q_m = -6.85$

Observed, En 54a.

Na^{22}

(Fig. 2)

I. $Na^{22}(\beta^+)Ne^{22}$ $Q_m = 2.840$

Most accurate measurement of half-life yields 2.58 ± 0.03 yr (Me 57). Also En 54a.

Decay proceeds predominantly by β^+ emission followed by a γ ray. Best magnetic spectrometer determinations of the β^+ end point are 542 ± 5 kev (Ma 50a) and 540 ± 5 kev (Wr 53), and of γ-ray energy, 1.277 ± 0.004 Mev (Al 49). Also En 54a.

Ground-state β^+ transition very weak. End point of this high-energy β^+ spectrum is 1.83 ± 0.06 Mev and intensity is (0.062 ± 0.015) percent (Wr 53). Log $ft = 13.1$. Also Mo 49.

Electron capture also occurs. Most accurate measurements yield: $EC/\beta^+ = (11.0 \pm 0.6) \times 10^{-2}$ (Sh 54b), $(12.3 \pm 1.0) \times 10^{-2}$ (Kr 54) and $(12.2 \pm 1.0) \times 10^{-2}$ (Al 55). Also Ho 53b, Ma 54c, Se 54, Ch 55, Di 55. Theory, with screening taking into account, yields 11.1×10^{-2} (Zw 54).

Internal conversion coefficient of 1.28-Mev γ ray measured as $(6.7 \pm 0.7) \times 10^{-6}$, compares with theoretical value for an E2 transition of 6.59×10^{-6} (Le 54). Spin of 1.28-Mev level in Ne^{22} is then 2^+. Spin of Na^{22} being $J = 3$ (Ma 50), with almost certainly even parity, predominant β^+ branch is allowed in agreement with spectrum shape (Ma 50a, Al 50b, Wo 54). Log $ft = 7.6$. Allowed character confirmed by absence of β-γ angular correlation (St 51a).

Theoretical discussion of Na^{22} ground-state spin, De 53a.

II. $F^{19}(\alpha,n)Na^{22}$ $Q_m=-1.928$

With α particles from an electrostatic generator, thresholds in neutron production observed at $E_\alpha=2.33$ and 3.04 Mev. At $E_\alpha>3.04$ Mev a γ ray observed of $E_\gamma=592\pm3$ kev, measured with a scintillation spectrometer. Half-life of corresponding Na^{22} level is <0.01 sec (He 54b).

Threshold, $E_\alpha=2.486$ Mev (Bu 56c) but corresponding Q (-2.061 Mev) would be in definite disagreement with Q from masses.

From threshold measurements and from neutron energy measurements with a proton recoil proportional counter, ground-state $Q=-2.0\pm0.2$ Mev and Na^{22} levels found at 0.45, 1.15, (1.8), 2.25, and 3.0 Mev, all ±0.2 Mev (Qu 56).

Resonances, see Na^{23}.

III. $Ne^{20}(d,p)Ne^{21}$ $E_b=11.274$ $Q_m=4.532$

Resonances observed corresponding to Na^{22} levels at 12.07 and 12.13 Mev (Go 55).

IV. $Ne^{21}(p,\gamma)Na^{22}$ $Q_m=6.742$

One resonance in γ-ray yield from a thin enriched Ne^{21} target bombarded with 0.6–1.3-Mev protons observed at $E_p=765$ kev (Br 47) corresponding to a Na^{22} level at 7.472 Mev.

V. $Ne^{21}(d,n)Na^{22}$ $Q_m=4.517$

Observed, En 54a.

VI. $Ne^{22}(p,n)Na^{22}$ $Q_m=-3.623$

Not observed.

VII. $Mg^{22}(\beta^+)Na^{22}$. See Mg^{22}

VIII. $Na^{23}(\gamma,n)Na^{22}$ $Q_m=-12.416$

Threshold measurements and yield curves, En 54a. Search for isomeric states in Na^{22} with half-lives in 10^{-5} sec–10^{-1} sec region unsuccessful (Ve 56).

IX. $Mg^{24}(d,\alpha)Na^{22}$ $Q_m=1.967$

Thin natural magnesium targets bombarded with 5-, 6-, and 7-Mev deuterons. From magnetic analysis ($\theta=90°$), the ground-state $Q=1.953\pm0.012$ Mev. Excited states at 0.585, 0.893, 1.535, 1.944, 1.990, 2.222, 2.579, 2.977, 3.067, and (3.525) Mev, all ±0.006 Mev. Upper limit of 3% of ground-state group given for intensity of any group which was not observed at $E_d=7$ Mev. Observed states should have $T=0$ from isobaric spin selection arguments (Br 55, Br 55a).

Also En 54a.

X. $Mg^{25}(p,\alpha)Na^{22}$ $Q_m=-3.140$

Enriched magnesium targets bombarded with protons from a 8-Mev electrostatic generator. From magnetic

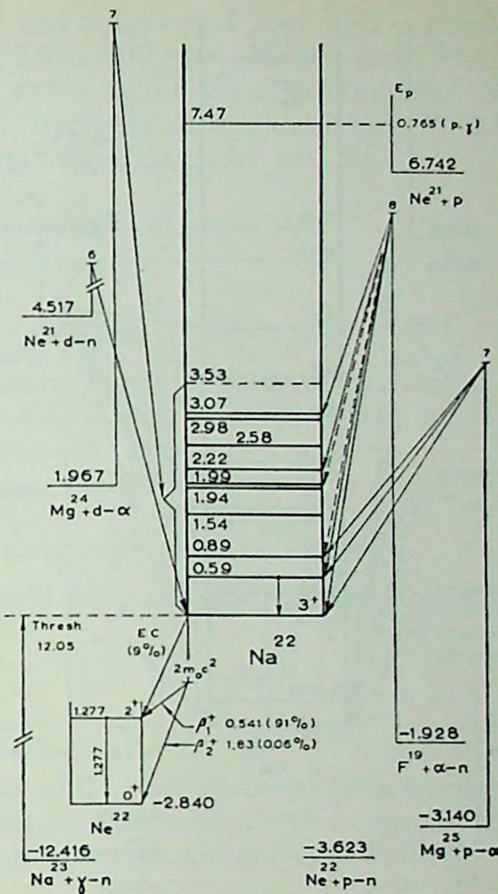


FIG. 2.

analysis ground-state $Q=-3.15$ Mev. Transitions observed to 0.59- and 0.89-Mev levels and some evidence for a level between those two (Br 55a).

Also Ba 54, Co 54a.

GENERAL REMARKS

The 0.59-Mev level probably not the $J=0^+$ $T=1$ level, expected at about 0.6 Mev (St 53b, St 53a, Mo 54, He 54b, Ta 54, Wi 56a), since this level is found from the $Mg^{24}(d,\alpha)Na^{22}$ reaction. The $T=1$ level may have been observed from the $Mg^{25}(p,\alpha)Na^{22}$ reaction.

Theoretical discussion of Na^{22} ground-state spin, Hi 54.

Mg^{22}

(not illustrated)

A 0.13-sec half-life observed from bombardment of natural magnesium with 23-Mev protons. This might be assigned either to Mg^{22} or to Al^{23} (Ty 54).

Na^{23}

(Fig. 3)

I. (a) $F^{19}(\alpha,n)Na^{22}$ $E_b=10.488$ $Q_m=-1.928$
(b) $F^{19}(\alpha,n\gamma)Na^{22}$ $E_b=10.448$ $E_x=0.592$
(c) $F^{19}(\alpha,p)Ne^{22}$ $E_b=10.448$ $Q_m=1.695$
(d) $F^{19}(\alpha,p\gamma)Ne^{22}$ $E_b=10.448$ $E_x=1.277$
(e) $F^{19}(\alpha,\alpha)F^{19}$ $E_b=10.448$
(f) $F^{19}(\alpha,\alpha\gamma)F^{19}$ $E_b=10.448$ $E_x=0.110$ and 0.197

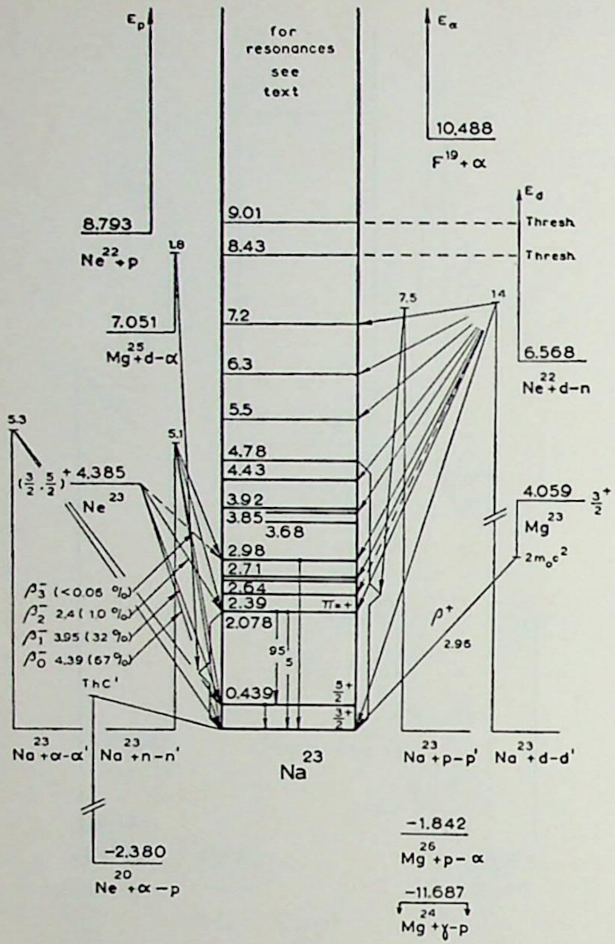


FIG. 3.

Yield of 592-kev radiation from reaction (b), of 1.28-Mev radiation from reaction (d), and of 110- and 197-kev radiation from reaction (f) measured in $E_\alpha=1.0$–3.5-Mev region. Relative yields are given at 18 resonances, all above 1.4 Mev (He 54b). Reactions (d) and (f) explored with better resolution in the $E_\alpha=0.6$–2.8-Mev region, yielding 26 resonances, all above 1.3 Mev (Sh 54e). Yield of reaction (f) at low E_α in agreement with Coulomb excitation theory for E2 excitation (He 54b, Sh 54c, Jo 54). Also, En 54a.

II. $Ne^{20}(\alpha,p)Na^{23}$ $Q_m=-2.380$

En 54a. Resonances, Mg^{24}.

III. (a) $Ne^{22}(p,\gamma)Na^{23}$ $E_b=8.793$
(b) $Ne^{22}(p,p)Ne^{22}$ $E_b=8.793$
(c) $Ne^{22}(p,p'\gamma)Ne^{22}$ $E_b=8.793$

Resonances, En 54a.

IV. $Ne^{22}(d,n)Na^{23}$ $Q_m=6.568$

Two slow-neutron thresholds observed from deuteron bombardment of natural neon with $Q=-1.866\pm0.010$ and -2.443 ± 0.015 Mev, corresponding to Na^{23} levels at $E_x=8.434\pm0.021$ and 9.012 ± 0.024 Mev (Ma 56b). Assignment based on observed intensities. Also Fr 55.

V. $Ne^{23}(\beta^-)Na^{23}$ $Q_m=4.385$

Half-life, 40.2 ± 0.4 sec (Br 50) and 37.6 ± 0.1 sec (Pe 57). Also En 54a, Ha 51a. Decay is complex. End point of ground-state transition ($67\pm3\%$ intensity) measured by scintillation spectrometer as 4.39 ± 0.05 Mev (Pe 57), in much better agreement with Q_m than a former magnetic spectrometer measurement (En 54a). A $32\pm3\%$ branch with an end point of 3.95 ± 0.05 Mev is in coincidence with 436 ± 4-kev γ rays, and a $1.00\pm0.15\%$ branch with 2.4 ± 0.1-Mev end point is in coincidence with 1.647 ± 0.016-Mev γ rays. The two γ rays are also coincident, and have intensities of 100: (3.0 ± 0.3). Higher energy γ rays present but have at most an intensity of 0.2% per disintegration. Beta transitions to three lowest Na^{23} states are evidently allowed with log $ft=5.25$, 5.38, and 5.88 (Pe 57). No superallowed branch occurs in Ne^{23} decay as formerly believed (Ki 55, Fe 55).

Results mentioned in the foregoing confirmed in Ge 55, Ge 56. End point of ground-state beta transition, 3.9 ± 0.3 Mev, lowest-energy γ ray, 440 ± 5 kev. A 3.0-Mev γ ray may be present but intensity relative to 440-kev γ ray is $<1\%$. Also Go 56.

Allowed character of beta transitions to $3/2^+$ Na^{23} ground state and $5/2^+$ first excited state limit Ne^{23} spin to $3/2^+$ or $5/2^+$. Same possibilities follow from a $Ne^{22}(d,p)Ne^{23}$ angular distribution measurement (Bu 56d).

VI. $Na^{23}(n,n'\gamma)Na^{23}$

In $E_n=440$–800-kev region, cross section for production of 0.44-Mev γ ray shows resonances at 542, 602, 633, 710, and 780 kev. The γ-ray angular distribution isotropic at 602- and 710-kev resonances, but anisotropic at 633 and 780 kev (Ha 56). At $E_n=2.5$ Mev, γ rays observed of 0.45 ± 0.01, 1.69 ± 0.03, and 2.2 ± 0.1 Mev (Wo 56). At $E_n=5.1$ Mev, γ rays reported of 0.439, 0.650, 1.63, 2.07, and 3.01 Mev (Mo 56c). The 0.650-Mev γ ray cannot easily be fitted into the Na^{23} level scheme. Also Cr 56a.

VII. (a) $Na^{23}(p,p')Na^{23}$
(b) $Na^{23}(p,p'\gamma)Na^{23}$

From magnetic analysis at several proton energies in the 7.0–7.5-Mev region and at several different angles, levels in Na^{23} observed at 0.440 ± 0.003, 2.078 ± 0.004, 2.393 ± 0.007, 2.641 ± 0.007, 2.705 ± 0.007, 2.983 ± 0.007, 3.678 ± 0.007, 3.850 ± 0.008, 3.915 ± 0.010, 4.431 ± 0.010, and 4.778 ± 0.010 Mev (Bu 57). Another value for first level, by magnetic analysis, is 437 ± 5 kev (Sc 56c). Most accurate determination of this level, from electrostatic analysis, is 439 ± 1 kev (Do 53). Also En 54a.

The 0.44-Mev γ-ray transition investigated extensively at $E_p=1.29$-Mev resonance. Internal conversion coefficient measured as $(4.9\pm0.6)\times10^{-5}$ in agreement with theory for M1 or E1 (Be 56a). Using both

$Na^{23}(p,p'\gamma)$ and $Na^{23}(\alpha,\alpha'\gamma)$ reactions, half-life of 0.44-Mev level determined with recoil techniques as between 10^{-13} and 10^{-12} sec (Kr 56). Also Sw 56. Gamma-ray angular distribution measurements at 1.29- and 1.46-Mev resonances limit spin of 0.44-Mev level to $5/2^+$ or $3/2^+$ (Re 56b, Be 56a).

Even parity of this level follows also from measurement of ratio of γ-ray yield from $Na^{23}(p,p'\gamma)$ and $Na^{23}(\alpha,\alpha'\gamma)$ reactions which is explained by E2 Coulomb excitation. From yield, partial half-life for E2 decay is computed as 3.0×10^{-10} sec (Te 56).

Second level in Na^{23} is fed at the $E_p=2.89$-Mev resonance. This level de-excites both to level (1) and to ground state, branching ratio of about 20. Ground-state transition is anisotropic, excluding $J=1/2$ for level (2) (Go 57b).

Other $Na^{23}+p$ resonances under Mg^{24}.

VIII. $Na^{23}(d,d')Na^{23}$

At $E_d=14$ Mev, levels in Na^{23} found at 1.9, 2.6, 3.75, 4.45, 5.5, 6.3, and 7.2 Mev (Bo 50). Also Ha 54b, El 56a. At $E_d=8$ Mev, groups proceeding to levels at 2.08, 2.71, and 3.85 Mev all show angular distributions compatible with $l_d=2$, which would determine parity of these levels as even (El 56a).

IX. $Na^{23}(\alpha,\alpha'\gamma)Na^{23}$

A 0.44-Mev γ ray observed (Te 54, Te 56, Kr 56). Half-life, quoted under reaction VII, Kr 56. The γ-ray angular distribution measured at $E_\alpha=2.5$ Mev is only compatible with $J=5/2^+$ for level (1) (Te 56).

X. $Na^{23}(N^{14},N^{14\prime}\gamma)Na^{23}$

A 435-kev γ ray observed resulting from Coulomb excitation by 15.6-Mev triply charged N^{14} ions (Al 56).

XI. $Mg^{23}(\beta^+)Na^{23}$ See Mg^{23}

XII. $Mg^{24}(\gamma,p)Na^{23}$ $Q_m=-11.687$

Cross section, Jo 55, Mo 55.

XIII. $Mg^{25}(d,\alpha)Na^{23}$ $Q_m=7.051$

Ground-state Q value, measured by magnetic analysis as 7.019 ± 0.013 Mev, differs from value computed from other links between Na^{23} and Mg^{25}. The Na^{23} levels at 0.44 and 2.08 Mev observed from this reaction (En 52).

XIV. $Mg^{26}(p,\alpha)Na^{23}$ $Q_m=-1.842$

Not observed.

Mg^{23}

(not illustrated)

I. $Mg^{23}(\beta^+)Na^{23}$ $Q_m=4.059$

Half-life, averaged from older determinations (En 54a) and from a more recent one (10.7 ± 0.7 sec; Hu 54), 11.9 ± 0.2 sec.

The β^+ end point$=2.82$ Mev with a cloud chamber (Wh 39), 2.99 ± 0.09 Mev (Bo 51), and 2.95 ± 0.07 Mev (Hu 54) with a scintillation spectrometer. No γ rays observed. Decay is superallowed ($\log ft=3.7$), determining Mg^{23} spin and parity as $3/2^+$.

II. $Ne^{20}(\alpha,n)Mg^{23}$ $Q_m=-7.222$

Not observed.

III. $Na^{23}(p,n)Mg^{23}$ $Q_m=-4.841$

Revision of generator calibration changed old value for threshold ($E_p=5.091\pm0.010$ Mev; Wi 52) to 5.053 ± 0.010 Mev (Ki 55a).

Resonances under Mg^{24}.

IV. $Mg^{24}(\gamma,n)Mg^{23}$ $Q_m=-16.529$

Threshold and yield, En 54a, Ka 54, Na 55, Mo 55, Ye 56.

Al^{23}

(not illustrated)

Half-life of 0.13 sec observed from bombardment of natural magnesium with 23-Mev protons. This might be assigned to either Al^{23} or Mg^{22} (Ty 54).

Na^{24}

(Fig. 4)

I. (a) $Na^{24}(\beta^-)Mg^{24}$ $Q_m=5.514$

Half-life, averaged from six older determinations (En 54a) and from a more recent measurement (14.90 ± 0.05 hr; To 55c), 15.00 ± 0.02 hr.

Decay scheme of Na^{24} investigated very thoroughly. The β^- decay goes predominantly to 4.12-Mev level of Mg^{24} followed by two γ rays in cascade through 1.37-Mev level. Accurate measurement of the β^- end point: 1.390 ± 0.005 Mev (Si 46). Best magnetic spectrometer determination of γ-ray energies is 2.7535 ± 0.0010 and 1.2680 ± 0.0010 Mev (He 52). Also En 54a. Log ft (for main β^- transition) 6.1.

Experimental material leading to an assignment of spins 0^+, 2^+, and 4^+ to Mg^{24} ground state and levels at 1.37 and 4.12 Mev, respectively, and of spin 4^+ to Na^{24} ground state, En 54a.

Ultra-fast delayed β-γ coincidence measurement of half-life of 1.37-Mev level, yielding $(2.5\pm1.7)\times10^{-11}$ sec, confirmed E2 character of 1.37-Mev transition (Co 55c).

A weak high-energy β^- component, proceeding to 1.37-Mev level, also found, with end point of 4.17 Mev and intensity of 0.003% ($\log ft=12.7$) (Tu 51). Also Gr 50a.

A 4.12-Mev crossover γ ray not definitely established (En 54a).

From $Na^{23}(p,\gamma)Mg^{24}$ measurements, a $J=3$ level at 5.23 ± 0.04 Mev found in Mg^{24}, de-exciting to 1.37 Mev

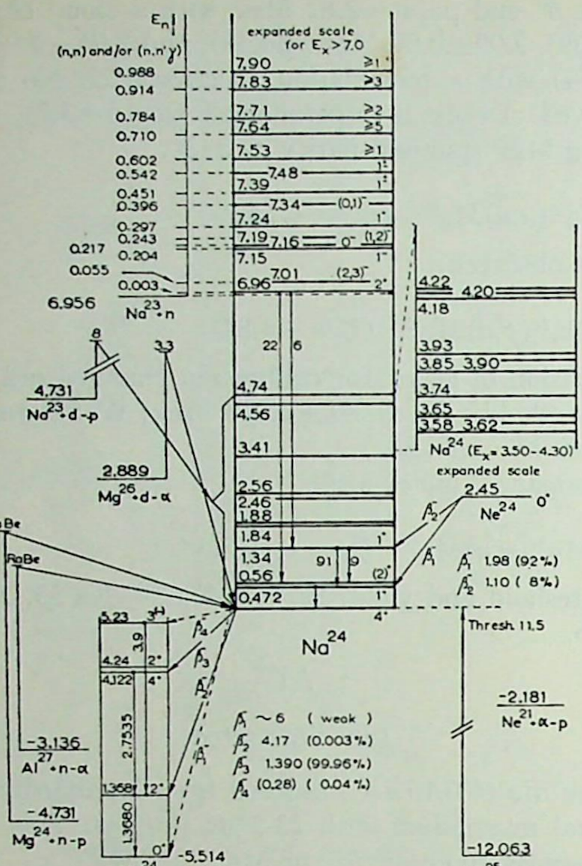


FIG. 4.

level with a 3.86 ± 0.04 Mev γ ray (Ne 54a). This level is probably also fed in $Na^{24}(\beta^-)Mg^{24}$ decay. A γ ray of 3.62 ± 0.05 Mev and intensity 0.04% has been reported (Be 51), while others report γ rays of 3.9 and 5.3 Mev (To 55b). Existence of the latter γ seems doubtful as its octupole character would exclude competition with 3.9-Mev dipole transition.

Log $ft=7.1$ of β^- decay to 5.23-Mev level points to a once-forbidden transition, yielding odd parity for 5.23-Mev level.

Theoretical discussion of Na^{24} spin, De 53a, Hi 54, Sc 54c.

(b) $Na^{24m}(\beta^-)Mg^{24}$ See $Ne^{24}(\beta^-)Na^{24}$ (reaction VII).

II. $Ne^{21}(\alpha,p)Na^{24}$ $Q_m=-2.181$

Not observed.

III. $Ne^{22}(d,\gamma)Na^{24}$ $Q_m=13.424$

Not observed (Al 55b).

IV. $Na^{23}(n,\gamma)Na^{24}$ $Q_m=6.956$

Cross-section measurements, Hu 55c, La 57.

Energies and intensities of thermal neutron capture γ rays determined by magnetic Compton spectrometer (Gr 55a), by magnetic pair spectrometer (Ki 51), with a two-crystal scintillation spectrometer (Br 56e) and with a magnetic lens spectrometer (Mo 56a) in Table II. Also Mi 50a. The 6.96-Mev ground-state transition not observed; its intensity is smaller than 0.1 per 100 captures (Gr 55a).

The intense γ ray A feeds the 0.56-Mev level, de-excited through the 0.472-Mev level by cascade $Z-Y$. Notable is absence of a 0.56-Mev γ ray. Upper limit for intensity is 3 per 100 captures (Gr 55a).

Few of the γ rays in Table II can be fitted in Na^{24} level scheme with certainty. For γ rays A, Y, and Z see above. Gamma-ray B feeds the 1.34-Mev level, which is de-excited by V to 0.47-Mev level, and probably by W to 0.56-Mev level. The 1.34-Mev level is not de-excited by U to Na^{24} ground state as this would conflict with the spin assignments found from the $Ne^{24}(\beta^-)Na^{24}$ decay (reaction VII). The γ rays N and P might feed the 1.34-Mev level from 3.85- and 3.58-Mev levels, respectively (Mo 56a).

IV. $Na^{23}(n,n)Na^{23}$ $E_b=6.956$

Cross sections and resonances, Hu 55c. Sharp resolved resonances at $E_n=2.9$ kev ($J=2^+$), 55 kev (3^- or 2^-; Bl 56, Cr 57), 204 kev (1^-), 217 kev (0^-), 243 kev (1^- or 2^-), 297 kev (1^+), 396 kev (0^- or 1^-), 451 kev (1^+ or 1^-), 542 kev (1^+), 602 kev (≥1), 710 kev (≥5), 784 kev (≥2), 914 kev (≥3), and 988 kev (≥1). Also Po 54, Pa 55, Jo 55a, To 55a, La 56.

γ rays from inelastic scattering under Na^{23}.

TABLE II. Thermal neutron capture γ rays from sodium (intensities in photons per 100 captures).

	Gr 55a		Ki 51		Mo 56a	
γ ray	Energy (Mev)	Intensity	Energy (Mev)	Intensity	Energy (Mev)	Intensity
A	6.40 ± 0.03	22	6.41 ± 0.03	20		
B	5.61 ± 0.03	6	5.61 ± 0.03	7.5		
B'	(5.30 ± 0.05)	0.3				
C	(5.12 ± 0.05)	0.8	5.13 ± 0.03	1.8		
C'	(4.90 ± 0.05)	1.2				
C''	(4.70 ± 0.05)	0.9				
D	4.50 ± 0.04	2.1				
D'	(4.30 ± 0.05)	0.5				
E	4.18 ± 0.04	2.1				
F	3.98 ± 0.04	17.2	3.96 ± 0.03	20		
G	3.86 ± 0.04	5.9	3.85 ± 0.03	11		
G'	(3.68 ± 0.05)	1.3				
H	3.60 ± 0.03	}18	3.60 ± 0.03	10	3.590 ± 0.025	18
I	3.56 ± 0.03		3.56 ± 0.03	20		
J	3.30 ± 0.02	5				
K	3.08 ± 0.02	9.5	Br 56e		3.070 ± 0.020	7
L	2.84 ± 0.02	7	Energy	Intensity		
M	2.68 ± 0.02	8.5	(Mev)			
N	2.52 ± 0.02	21	2.53 ± 0.03	19	2.510 ± 0.020	~15
O	2.41 ± 0.03	10.5				
P	2.21 ± 0.03	7.5			2.210 ± 0.015	8
Q	2.02 ± 0.015	11.5	2.02 ± 0.03	12	2.030 ± 0.010	13
R	(1.95 ± 0.03)	4				
S	(1.87 ± 0.03)	5.5			1.900 ± 0.020	~3
T	1.66 ± 0.01	7.5	1.66 ± 0.05	5	1.630 ± 0.008	10
U	1.35 ± 0.01	6.5	1.35 ± 0.03	6		
V	0.86 ± 0.01	44	0.86 ± 0.02	34	0.877 ± 0.005	30 ± 10
W	0.79 ± 0.015	4.3				
X	(0.71 ± 0.015)	5				
Y	0.47 ± 0.015	74	0.48 ± 0.02	60	0.473 ± 0.004	50 ± 10
Z					0.090 [a]	

[a] Observed with a scintillation spectrometer (Mo 56a).

V. $Na^{23}(n,p)Ne^{23}$ $E_b=6.956$ $Q_m=-3.603$

Cross section from threshold to $E_n=6.0$ Mev; resonances observed corresponding to Na^{24} levels at $E_x=11.93$ and 12.33 Mev (Bo 57a).

VI. $Na^{23}(d,p)Na^{24}$ $Q_m=4.731$

Ground-state Q measured by magnetic analysis $=4.731\pm0.007$ Mev (Sp 52) and 4.723 ± 0.008 Mev (Mi 52). Eighteen levels in Na^{24} found by high-resolution magnetic analysis at $\theta=90°$ and at deuteron energies between 1.5 and 2.2 Mev (Sp 52) in Table III. Also En 54a, El 56a.

The l_n values found by comparison with stripping theory in Table III. Brackets indicate unresolved groups.

VII. $Ne^{24}(\beta^-)Na^{24}$ $Q_m=2.45$

New radioactive nuclide Ne^{24} recently produced (Dr 56) from $Ne^{22}(t,p)Ne^{24}$ reaction. Half-life, 3.38 ±0.02 min.

Two (noncoincident) γ rays observed with a scintillation spectrometer with following energies and relative intensities: γ_1, 472 ± 5 kev (100); γ_2, 878 ± 9 kev (8 ± 2). Upper limit of 0.5% given for potential 92- and 564-kev γ rays, and of 1% for a 785-kev γ ray. Two β^- ray branches found with a plastic scintillator with following end points and relative intensities: β_1^- 1.98 ± 0.05 Mev (92 ± 2), β_2^- 1.10 ± 0.05 Mev (8 ± 2). The β_1^- branch not in coincidence with γ rays, the β_2^- branch coincident with γ_2. Both β^- transitions are allowed ($\log ft$ $=4.4$ for both) and have linear Fermi plots. A weak β^- branch with ~6-Mev end point also may be present.

The 472-kev level in Na^{24} is isomeric. From decay of γ_1, half-life estimated as 20^{+30}_{-15} msec. This explains absence of coincidences of γ_1 with γ_2 or with β^- particles. The weak high-energy β^- branch may be explained as a transition from isomeric state to Mg^{24} ground state.

Observed half-life of 472-kev transition uniquely establishes it as octupole. Spin of 472-kev level must then be $J=1^+$, because of allowed character of β^- transition to this level. Also 1.34-Mev level has $J=1^+$; from allowed character of corresponding β^- transition follows $J=0^+$ or 1^+, while $Na^{23}(d,p)Na^{24}$ angular distribution measurements (see Reaction VI) lead to $J=1^+$ or 2^+.

That the 0.56-Mev level is not fed in the Ne^{24} β^- decay may be explained by spin assignments of $J=0^+$ or $J\gtrsim2$. A $0^+\rightarrow0^+$ β^- transition would be forbidden by isobaric spin selection rule. Spin 1^+ or 2^+ follows from $Na^{23}(d,p)Na^{24}$ angular distribution measurements (see Reaction VI). The 0.56-Mev level then has most likely $J=2^+$, in agreement with fact that no 0.56-Mev γ ray is observed (also Reaction III).

TABLE III. Levels in Na^{24} from $Na^{23}(d,p)Na^{24}$ and corresponding neutron orbital momenta l_n.

Reference: Group	Sp 52 Q value (Mev)	Sp 52 Na^{24} level (Mev)	Ta 53 [a] l_n	Sh 54 [b] l_n	Br 54c [c] l_n	El 56 [d] l_n
(0)	4.731 ± 0.007	0	2	2	2	2
(1)	4.259 ± 0.007	0.472 ± 0.008		}0+2	}0+2	2
(2)	4.167 ± 0.007	0.564 ± 0.008				0
(3)	3.390 ± 0.006	1.341 ± 0.008		0	0	0
(4)	2.887 ± 0.006	1.844 ± 0.008				}0+2
(5)	2.847 ± 0.006	1.884 ± 0.008				
(6)	2.267 ± 0.006	2.464 ± 0.008				}1+2
(7)	2.170 ± 0.006	2.561 ± 0.008				
(8)	1.322 ± 0.005	3.409 ± 0.008				1
(9)	1.149 ± 0.006	3.582 ± 0.009				
(10)	1.108 ± 0.006	3.623 ± 0.009				
(11)	1.083 ± 0.006	3.648 ± 0.009				
(12)	0.993 ± 0.005	3.738 ± 0.008				
(13)	0.881 ± 0.005	3.850 ± 0.008				
(14)	0.832 ± 0.005	3.899 ± 0.008				
(15)	0.802 ± 0.005	3.929 ± 0.008				1
(16)	0.547 ± 0.005	4.184 ± 0.008				}1
(17)	0.529 ± 0.005	4.202 ± 0.008				
(18)	0.512 ± 0.005	4.219 ± 0.008				
(19)	0.173 ± 0.005	4.558 ± 0.009				1
(20)		4.74 ± 0.07 [e]				1

[a] $E_d=1.15$ Mev; nuclear emulsions.
[b] $E_d=3$ Mev; Al absorption.
[c] $E_d=10$ Mev; nuclear emulsions.
[d] $E_d=8$ Mev; magnetic analysis. Groups (1) and (2) were only partly resolved in this investigation.
[e] El 56a.

VIII. $Mg^{24}(n,p)Na^{24}$ $Q_m=-4.731$

Cross-section data, En 54a, Co 56, Ve 57.

IX. $Mg^{25}(\gamma,p)Na^{24}$ $Q_m=-12.063$

Threshold and cross-section data, En 54a, Ka 54, Mo 55.

X. $Mg^{26}(d,\alpha)Na^{24}$ $Q_m=2.889$

Cross section, En 54a.

XI. $Al^{27}(n,\alpha)Na^{24}$ $Q_m=-3.136$

Cross section, En 54a.

Mg^{24}

(Fig. 5)

I. $Ne^{20}(\alpha,p)Na^{23}$ $E_b=9.307$ $Q_m=-2.380$

Observed at $E_\alpha=3.923$-Mev resonance (Table IV) (Go 54c).

II. $Ne^{20}(\alpha,\alpha)Ne^{20}$ $E_b=9.307$

Thirteen sharp resonances observed in neon elastic cross section at four different scattering angles for $E_\alpha=2$–4 Mev. Eleven assigned to Ne^{20} (Table IV) and two to Ne^{22}. Spins and parities from partial wave analysis (Go 54c). Reaction IV gives comparison with $Na^{23}+p$ resonances.

III. $Ne^{21}(\alpha,n)Mg^{24}$ $Q_m=2.550$

Not observed. Reaction chain $Ne^{20}(p,\gamma)Na^{21}(\beta^+)$ $Ne^{21}(\alpha,n)Mg^{24}$ believed to be a source of neutrons in a certain stage of stellar development (Ho 56).

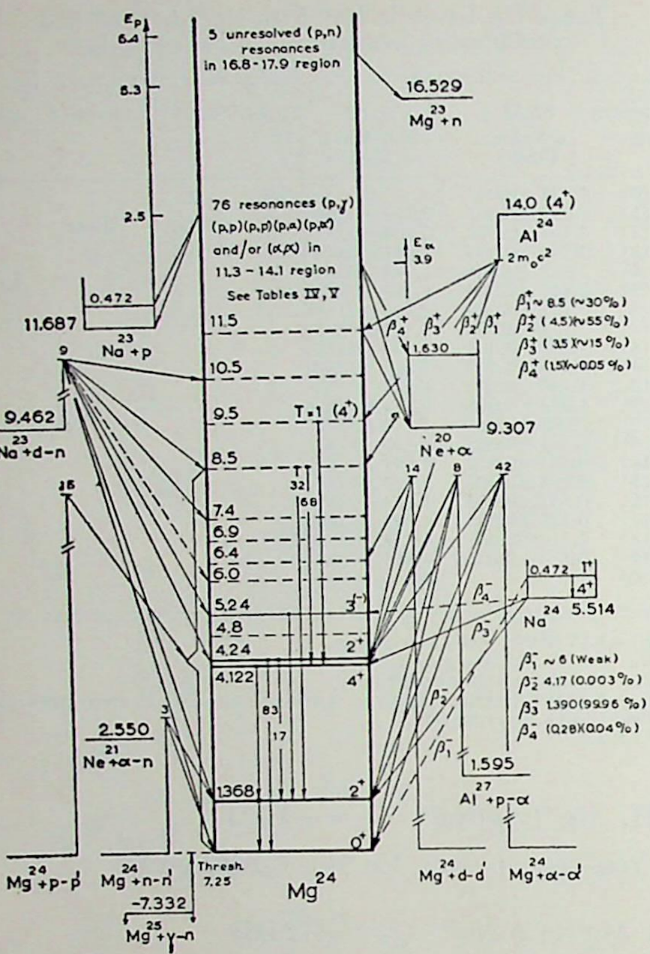


FIG. 5.

IV. (a) $Na^{23}(p,\gamma)Mg^{24}$ $E_b=11.687$
(b) $Na^{23}(p,n)Mg^{23}$ $E_b=11.687$ $Q_m=-4.842$
(c) $Na^{23}(p,p)Na^{23}$ $E_b=11.687$
(d) $Na^{23}(p,p'\gamma)Na^{23}$ $E_b=11.687$ $E_x=0.439$
(e) $Na^{23}(p,\alpha)Ne^{20}$ $E_b=11.687$ $Q_m=2.380$
(f) $Na^{23}(p,\alpha\gamma)Ne^{20}$ $E_b=11.687$ $E_x=1.630$

Large amount of information has become available in recent years on $Na^{23}+p$ resonance reactions (Table V, also En 54a). Most accurate measurements taken for proton resonant energies and total widths given in Table V. Besides references under Table V, also Ny 55 (energies of 5 resonances in 0.5–1.0-Mev region with relative γ-ray yields), Te 54a (energies of 14 resonances in 0.85–1.70-Mev region with widths and relative γ-ray yields), and Fa 55. Relative γ-ray yields of 5 resonances in 0.22–0.53-Mev region, Ha 55c. For absolute α_0 yields at the 287-, 338-, and 594-kev resonances, Fl 54.

Spin and parity assignments of 308-, 511-, 594-, and 675-kev resonances based on measurements of γ-ray spectra, angular distributions, γ-γ angular correlations, and γ-ray widths. The 4.24-Mev Mg^{24} level assigned $J=2^+$. Level at 5.26 Mev observed with $J=3$ and probably negative parity (Gr 55e). Also Ca 53, Tu 53, Ca 54a, Hi 54a, for measurements of γ-ray spectrum at 308-kev resonance.

Spin and parity assignments at 1012-, 1137-, 1166-, and 1288-kev resonances in St 54a based on α_0 angular distribution measurements, those in Ba 56 mainly on p_0 and some on γ angular distribution measurements, and those in Ne 54 mainly on partial widths for different decay processes. Gamma-ray angular distribution and correlation measurements at 1398-kev resonance lead also to assignments of $J=2$ to Mg^{24} 4.24-Mev level (decaying both to ground state and 1.37-Mev level with relative intensity of 1:0.2) and $J=3$ to a 5.23±0.04-Mev level (decaying through 1.37-Mev level) (Ne 54). The 1^+ assignment to 1258-kev resonance in Se 53 based on $\alpha_1-\gamma$ angular correlation measurements in planes parallel with and perpendicular to proton beam.

On the average, Mg^{24} excitation energies as found from $Ne^{20}+\alpha$ are 5 kev higher than those from $Na^{23}+p$. Except for this, there is beautiful agreement between excitation energies, spins and parities of Mg^{24} levels found from $Ne^{20}+\alpha$ elastic scattering (Table IV) and from $Na^{23}+p$ (Table V). In former reaction, only $T=0$ levels observed with "natural parity" (parity *even* if J is even, parity *odd* if J is odd).

The 12.256-Mev level must be a doublet, Gr 55e. A unique 2^- assignment follows from $Na^{23}(p,\gamma)$, in disagreement with 3^- assignment from $Na^{23}(p,p)$ and $Ne^{20}(\alpha,\alpha)$. Negative parity follows also from a polarization measurement of 10.8-Mev capture radiation, Hu 56.

TABLE IV. Resonances in $Ne^{20}(\alpha,\alpha)Ne^{20}$ reaction and corresponding Mg^{24} levels (Go 54c).

E_α (lab) (Mev)	E_x (Mg^{24}) (Mev)	Γ (kev)	J,π
2.488	11.380	(0.5)	1^-
2.573	11.451	(1)	0^+
2.652	11.517	(0.5)	2^+
2.903	11.726	10±2	0^+
3.062	11.858	8±2	1^-
3.184	11.960	(0.5)	2^+
3.548	12.263	(1)	3^-
3.780	12.456	7±2	1^-
3.801	12.474	5±1	2^+
3.839	12.506	(1)	4^+
3.923	12.576	6±1	2^+
all ±0.005			

V. $Na^{23}(d,n)Mg^{24}$ $Q_m=9.462$

Angular distributions of neutrons, detected with nuclear emulsions, measured at $E_d=8$ Mev. Transitions to Mg^{24} ground state and 1.37-Mev level very weak; those to 4.12- and 4.24-Mev levels unresolved and weak ($l_p=0+2$); those to 5.1-, 5.5-, and 6.3-Mev levels weak and isotropic. Transitions observed to levels at 7.45, 8.48, and 10.46 Mev, all ±0.10 Mev, with $l_p=0$, $0+2$, and 0, respectively, El 56a.

Essentially same results obtained at $E_d=9$ Mev with a proton recoil triple ionization chamber for neutron detection. Transitions to a 7.5-Mev level assigned $l_p=0+2$, those to a 8.4-Mev level $l_p=0$. No higher levels observed, Ca 55. Also En 54a.

VI. $Na^{24}(\beta^-)Mg^{24}$ See Na^{24}.

VII. $Mg^{24}(e,e')Mg^{24}$

Differential cross section for elastic scattering of 187-Mev electrons and for inelastic scattering leading to level (1) determined, (He 56).

VIII. $Mg^{24}(n,n')Mg^{24}$

At $E_n=2.56$ Mev, only one γ ray ($E_\gamma=1.368\pm0.010$ Mev) from inelastic neutron scattering on natural magnesium can be assigned to Mg^{24} (Da 56c). Also En 54a, Sc 54d, Pa 55, Ra 55, Cr 56a. Elastic neutron scattering, Cu 56.

IX. $Mg^{24}(p,p')Mg^{24}$

Most accurate measurement of first Mg^{24} level, done by electrostatic analysis of inelastically scattered protons, 1.371 ± 0.002 Mev (Do 53). Second and third levels measured by magnetic analysis at 4.13 ± 0.02 and 4.24 ± 0.02 Mev (1/a 52). Higher levels reported (all from natural magnesium bombardments) at 4.77, 5.24, 6.00, 6.42, and 6.90 Mev, all ±0.05 Mev (Gr 56f), at 5.10 Mev (St 52), at 5.1 ± 0.1, 5.9 ± 0.1, and 6.3 ± 0.1 Mev (Fi 54), at 5.51 ± 0.3, 7.32 ± 0.3, and 8.30 ± 0.4 Mev (Fu 48), and at 6.38 ± 0.08 Mev (Ba 52).

Angular distribution measurements of elastically and inelastically scattered protons, En 54a, Fi 54, Gr 55, Co 55, Gr 56f, Gu 56a, Sh 56.

A p-γ angular correlation measurement, En 54a.

Resonances, under Al^{25}.

X. $Mg^{24}(d,d')Mg^{24}$

Angular distribution measurements of elastically and inelastically scattered deuterons, En 54a, Gr 55, Ha 56c.

XI. $Mg^{24}(\alpha,\alpha')Mg^{24}$

Angular distribution measurements of elastically and inelastically scattered α particles, Gu 56, Wa 56a. Groups both to 1.37-Mev level and to doublet at 4.2 Mev agree with direct surface interaction for $0^+\to2^+$ transitions. Thus, the 2^+ 4.24-Mev level is excited but not the 4^+ 4.12-Mev level (Wa 56a).

Resonances, under Si^{28}.

XII. $Al^{24}(\beta^+)Mg^{24}$ See Al^{24}.

XIII. $Mg^{25}(\gamma,n)Mg^{24}$ $Q_m=-7.332$

Threshold and yield, En 54a, Ka 54, Na 55, Ye 56, Mo 55.

XIV. $Al^{27}(\gamma,t)Mg^{24}$ $Q_m=-18.217$

Threshold, 18.2 ± 0.2 Mev. Yield from 31-Mev bremsstrahlung in good agreement with statistical theory (Wa 56).

TABLE V. Resonances in $Na^{23}+p$ reactions.

E_p (lab) (kev)	E_x (Mg^{24}) (Mev)	Decay *					Γ (kev)	J, π
250.8 ±0.2 [a]	11.928					γ	0.3 ±0.2 [a]	
287 ±1.5 [b]	11.962			α_0				
307.8 ±0.3 [a]	11.983					γ	0.8 ±0.3 [a]	2^- [d]
338 ±1.5 [b]	12.011			α_0				
373.5 ±0.4 [a]	12.045					γ	2.0 ±1.0 [a]	
443.8 ±0.6 [a]	12.113					γ	0.8 ±0.3 [a]	
510.9 ±0.6 [a]	12.177					γ	0.8 ±0.3 [a]	1^+ [d]
594 [c]	12.256	p_0		α_0		γ	2 ±1 [c]	2^- [d], 3^- [c]
675 [c]	12.334	p_0				γ	≤1 [c]	3^+ [d]
725 [c]	12.382	p_0					7 ±2 [c]	0^- [c]
740 [c]	12.396	p_0				γ	<3 [c]	
744 [c]	12.400					γ	<3 [c]	
797 [c]	12.451	p_0		α_0			7 ±2 [f]	1^- [c]
815 [c]	12.468	p_0		α_0			5 ±1 [f]	2^+ [c]
849 [c]	12.501	p_0					(1) [f]	4^+ [c]
877 [c]	12.527	p_0				γ	8 ±2 [c]	1^+ [c], (1) [g], $(1,2)^+$ [j]
922 [c]	12.571	p_0		α_0			6 ±1 [f]	2^+ [c]
989 [c]	12.635					γ	≤1 [c]	4 [c]
1011 [c,e]	12.656		p_1			γ	≲0.5 [c]	
1012 [c,e]	12.657		p_1	α_0	α_1		0.8 ±0.1 [e]	3^- [h], (2^+) [g]
1022 [c,e]	12.666	p_0	p_1		α_1	γ	6.6 ±0.5 [e]	2^- [c]
1087 [c,e]	12.729		p_1			γ	1.1 ±0.5 [e]	
1094 [c,e]	12.734		p_1	α_0	α_1		7.9 ±0.5 [e]	(2^+) [g]
1137 [c,e]	12.777			α_0			30 ±5 [e]	0^+ [h]
1166 [c,e]	12.804		p_1	α_0	α_1	γ	1.2 ±0.1 [e]	2^+ [h], (1^-) [g]
1176 [c,e]	12.814	p_0	p_1		α_1	γ	2.5 ±0.5 [e]	1^+ [c]
1206 [c,e]	12.844	p_0	p_1		α_1		0.3 ±0.1 [e]	(2^-) [c]
1213 [c,e]	12.849	p_0	p_1		α_1	γ	0.4 ±0.1 [e]	(3^+) [c]
1258 [c,e]	12.893	p_0	p_1		α_1		0.6 ±0.1 [e]	1^+ [c,i]
1273 [c]	12.907					γ		
1288 [c,e]	12.921	p_0	p_1	α_0	α_1		7.1 ±0.2 [e]	1^- [c,h]
1321 [c,e]	12.953	p_0	p_1		α_1	γ	2.1 ±0.5 [e]	3^+ [c]
1329 [c,e]	12.962	p_0	p_1		α_1		3.5 ±0.2 [e]	
1334 [c,e]	12.966	p_0	p_1				6.7 ±0.2 [e]	
1365 [c,e]	12.996	p_0	p_1				1.2 ±0.2 [e]	(0^-) [c]
1398 [c,e]	13.028	p_0	p_1		α_1	γ	0.8 ±0.1 [e]	3^+ [c,g]
1419 [c,e]	13.048		p_1	α_0	α_1	γ	≤0.3 [e]	4^+ [c]
1460 [c,e]	13.087	p_0	p_1	α_0	α_1		9.8 ±0.2 [e]	3^- [c], (2^-) [g]
1511 [e]	13.135	... [k]		α_0		[l]	<10 [e]	
1519 [e]	13.143	...	p_1		α_1		3.2 ±0.5 [e]	
1558 [e]	13.170	...	p_1	α_0	α_1		6.5 ±0.5 [e]	
1575 [e]	13.196	...		α_0			<10 [e]	
1638 [e]	13.257	...	p_1	α_0	α_1		~50 [e]	
1645 [e]	13.263	...	p_1				~8 [e]	
1653 [e]	13.271	...			α_1		~2 [e]	
1719 [e]	13.334	...	p_1		α_1		2.0 ±0.5 [e]	
1737 [e]	13.352	...	p_1				15.2 ±2 [e]	
1749 [e]	13.363	...	p_1				≤1.0 [e]	
1802 [e]	13.414	...	p_1	α_0			4 ±1 [e]	
1805 [e]	13.417	...		α_0	α_1		6 ±1 [e]	
1832 [e]	13.443	...	p_1		α_1		<0.35 [e]	
1839 [e]	13.449	...	p_1		α_1		3.2 ±0.2 [e]	
1870 [e]	13.479	...	p_1				~1 [e]	
1933 [e]	13.540	...	p_1		α_1		6.9 ±0.5 [e]	
1976 [e]	13.581	...			α_1		23 ±3 [e]	
1979 [e]	13.584	...	p_1	α_0			9.5 ±1 [e]	
2027 [e]	13.629	...			α_1		3.3 ±1 [e]	
2075 [e]	13.676	...		α_0	α_1		8.7 ±0.5 [e]	
2080 [e]	13.681	...	p_1				31 ±2 [e]	
2122 [e]	13.721	...	p_1	α_0	α_1		4.8 ±0.5 [e]	
2170 [e]	13.767	...	p_1	α_0			~2 [e]	
2200 [e]	13.795	...	p_1	... [m]	α_1		3.5 ±0.5 [e]	
2223 [e]	13.817	...	p_1	...	α_1		47 ±5 [e]	
2243 [e]	13.836	...	p_1	...			~2 [e]	
2284 [e]	13.876	...	p_1	...	α_1		2.0 ±0.5 [e]	
2297 [e]	13.888	...	p_1	...	α_1		~28 [e]	
2340 [e]	13.929	...	p_1	...			~10 [e]	
2343 [e]	13.932	...		...	α_1		~14 [e]	
2354 [e]	13.943	...	p_1	...	α_1		~4 [e]	
2388 [e]	13.976	...	p_1	...	α_1		4.9 ±0.5 [e]	
2436 [e]	14.021	...	p_1	...	α_1		6.5 ±0.5 [e]	

* For observed modes of decay p_0 refers to elastic scattering, p_1 to inelastic protons (observed through the 0.439-Mev Na^{23} γ ray), α_0 to ground-state α particles, α_1 to short-range α particles (mostly observed through 1.630 Mev Ne^{20} γ rays), and γ to hard γ rays from proton capture.

[a] (Ha 55c); [b] (Fl 54); [c] (Ba 56, Pr 56); [d] (Gr 55e); [e] (St 54); [f] (Go 54c); [g] (Ne 54); [h] (St 54a); [i] (Se 53); [j] (De 56).

[k] No elastic proton scattering measurements available above $E_p=1.5$ Mev.

[l] Hard γ rays very weak or absent above $E_p=1.5$ Mev.

[m] No ground-state α-particle measurements available above $E_p=2.2$ Mev.

XV. $Al^{27}(p,\alpha)Mg^{24}$ $Q_m=1.595$

Q by electrostatic analysis $=1.594\pm0.002$ Mev (Do 53), and by magnetic analysis $=1.585\pm0.015$ Mev (Fr 50), 1.595 ± 0.007 Mev (Va 52a), 1.61 ± 0.02 Mev (Ru 53a), and 1.596 ± 0.006 Mev (Va 57d).

From electrostatic analysis first level obtained at

1.366±0.004 Mev (Do 53), and from magnetic analysis at 1.366±0.006 Mev (Va 57d).

Transitions to 4.11- and 4.21-Mev levels observed by magnetic analysis at $E_p=8$ Mev (Re 52).

Resonances, under Si^{28}. Also En 54a, Gr 54a, Pa 54.

Al^{24}

(not illustrated)

I. $Al^{24}(\beta^+)Mg^{24}$ $Q_m=14.0$

Half-life (average of three measurements, En 54a) 2.09±0.04 sec.

Gamma rays observed with a scintillation spectrometer in Table VI. Positrons reported with endpoint of ~8.5 Mev. One per several thousand disintegrations proceeds through an excited state in Mg^{24} which emits α particles of $E_\alpha \sim 2$ Mev (Gl 55).

The β^+ decay seems best explained by assuming a 30% branch to an 8.5-Mev level, probably with spin 3^+, de-exciting through the 7.12- and 4.22-Mev γ rays to the 1.37-Mev 2^+ and 4.12-Mev 4^+ levels, a 15% branch to a 9.5-Mev level, de-exciting through the 5.35-Mev γ ray to the 4.12-Mev level, and finally a weak transi-

TABLE VI. Gamma rays from the $Al^{24}(\beta^+)Mg^{24}$ decay.

Author:	Gl 55		Br 54a
	E_γ (Mev)	Rel. int.	E_γ (Mev)
	1.39±0.03	0.40	1.38±0.04
	2.73±0.06	0.32	2.70±0.06
	4.22±0.10	0.15	4.21±0.12
	5.35±0.10	0.06	5.37±0.14
			(5.66±0.18)
	7.12±0.10	0.07	7.02±0.20

ion to a level at ~11 Mev, de-exciting through α emison (Gl 55). The 9.5-Mev level in position expected ɔr $T_z=0$ member of $T=1$ triplet to which also belong $=4^+$ Na^{24} ground state and Al^{24} ground state (presumably also with spin $J=4^+$). The β^+ transition to this level would then be favored (Bo 55, Wi 56a).

II. $Mg^{24}(p,n)Al^{24}$ $Q_m=-14.8$

Threshold is 15.4±0.3 Mev (Bi 52). Also Ty 54, Co 55b.

Na^{25}

(not illustrated)

I. $Na^{25}(\beta^-)Mg^{25}$ $Q_m=4.0$

Measurements of half-life (En 54a, Iw 55, Na 56), in good agreement, average to 59.5±0.7 sec. The β^- end point, measured with Al absorbers, 3.7±0.3 Mev (Bl 47) and with a scintillation spectrometer, 4.0±0.2 Mev (Ma 55). From analysis of γ spectrum and β^--γ coincidence spectrum, 65%, 3.5%, 25%, and 6.5% of β^- transitions lead to ground state and to 0.58-, 0.98-, and 1.61-Mev levels of Mg^{25}, respectively (Ma 55). However, intensity ratios in γ spectrum measured with a large NaI crystal show that β^- branching to the 0.58-Mev level is probably less than 1% (Go 56). Branching ratios, log ft values, spins, and parities in Table VII. Assignment of $5/2^+$ to Na^{25} ground state based upon absence of a measurable transition to Mg^{25} (1) together with allowed character of transitions to Mg^{25} (0) and (2). Energies and relative intensities of γ rays in Table VIII. Discussion of possibility of favored transitions, Ki 55, Fe 55.

TABLE VII. The β^- decay of Na^{25} (Go 56).

Mg^{25} level (Mev)	0	0.58	0.98	1.16	1.96
Branching ratio (%)	65	<1	30	5	<1
Log ft	5.2	>6.9	5.2	5.6	>5.9
J and π; $Na^{25}(0)=5/2^+$	$5/2^+$	$1/2^+$	$3/2^+$	$(7/2)^+$	$5/2^+$

II. $Ne^{22}(\alpha,p)Na^{25}$ $Q_m=-3.70$

Observed, En 54a.

III. $Mg^{25}(n,p)Na^{25}$ $Q_m=-3.22$

Cross section, En 54a.

IV. $Mg^{26}(\gamma,p)Na^{25}$ $Q_m=-14.21$

Threshold, $E_\gamma=14.0\pm1.0$ Mev. Cross section, En 54a, Ka 54; also Mo 55.

TABLE VIII. Gamma rays following the β^- decay of Na^{25}.[a]

	Iw 55	Ma 56	Na 56	Go 56
γ_1	410	384±10 (44±4)	370±10 (108)	400 (95±4)
γ_2	590	576±10 (59±7)	580±10 (135)	580 (89±5)
γ_3	980	978±15 (100±15)	975 (100)	980 (100±3)
γ_4		1603±20 (40±7)		1610 (33±3)

[a] Energies in kev; relative intensities normalized to $\gamma_3=100$ and placed in brackets.

Mg^{25}

(Fig. 6)

I. $Ne^{22}(\alpha,n)Mg^{25}$ $Q_m=-0.483$

Work with Po α-particle sources, En 54a.

II. $Mg^{24}(n,\gamma)Mg^{25}$ $Q_m=7.332$

Thermal neutron capture cross section of natural magnesium is 63±4 mb; those of Mg^{24}, Mg^{25}, and Mg^{26} are 33±10, 270±90, and 60±60 mb, and their abundances 78.6, 10.1, and 11.3%, respectively, indicating approximately 41% of thermal neutron captures in natural magnesium should occur in Mg^{24} (Hu 55c).

Gamma rays from capture of thermal neutrons in natural magnesium, measured with magnetic pair spectrometer (Ki 51, Ki 53a) and with two-crystal scintillation spectrometer (Br 56e), in Table IX. Intensity of

line J not in agreement with above estimate of contribution of Mg^{24} to capturing cross section of natural magnesium. Possible assignments in Table IX based mainly upon comparison of γ energies with Q values of (d,p) reactions. Great intensity of line J (from capturing state to p level at 3.405 Mev in Mg^{25}) explained by assuming this is most energetic E1 transition. Since this level is de-excited by γ rays K and L to ground state (spin $5/2^+$) and first level (spin $1/2^+$), both these competing transitions must be E1, and spin of 3.405 level is uniquely determined as $3/2^-$ (Ki 54a).

Recalculation of intensities of some lines in photons per 100 captures in Mg^{24}, Ki 54a.

III. $Mg^{24}(n,n)Mg^{24}$ $E_b=7.332$

Total neutron cross section of natural magnesium, Hu 55c.

TABLE IX. Gamma rays from capture of thermal neutrons in natural magnesium.

γ ray	Energy in Mev	Intensity in photons per 100 captures in natural Mg	Final nucleus and probable transition
A''	11.086 ± 0.025 [d]		Mg^{26} $(c)\rightarrow(0)$
A'	10.08 ± 0.02 [d]		Mg^{26} $\rightarrow(0)$
A	9.26 ± 0.04 [a]	1 [a]	Mg^{26} $(c)\rightarrow(1)$
B''	8.93 ± 0.02 [d]		Mg^{26} $\rightarrow(0)$
B'	8.55 ± 0.02 [d]		Mg^{26} $\rightarrow(0)$
B	8.16 ± 0.03 [a]	9 [a]	Mg^{26} $(c)\rightarrow(2)$
C	7.37 ± 0.08 [a]	0.5 [a, e]	Mg^{25} $(c)\rightarrow(0)$
D	7.15 ± 0.04 [a]	1 [a]	Mg^{26} $(c)\rightarrow(3)$
E	6.75 ± 0.04 [a]	2.5 [a, e]	(Mg^{26}), Mg^{25} $(c)\rightarrow(1)$
F_1	6.440 ± 0.008 [b]	1.3 [b, e]	Mg^{27} $(c)\rightarrow(0)$
F_2	6.358 ± 0.007 [b]	3.5 [b, e]	Mg^{25} $(c)\rightarrow(2)$
G	5.73 ± 0.04 [a]	1 [a]	(Mg^{25}) $(c)\rightarrow(3)$
H	5.50 ± 0.04 [a]	7 [a]	(Mg^{25}), (Mg^{26})
I	5.05 ± 0.07 [a]	9 [a]	(Mg^{25})
J	3.918 ± 0.004 [b]	70 [b, e]	Mg^{25} $(c)\rightarrow(8)$
K	3.408 ± 0.018 [d]	$\sim$10 [d]	Mg^{25} $(8)\rightarrow(0)$
L	2.816 ± 0.016 [d]	$\sim$50 [a, c, d]	Mg^{25} $(8)\rightarrow(1)$
M	1.87 ± 0.03 [c]	22 [c]	Mg^{26} $(1)\rightarrow(0)$
N	1.07 ± 0.05 [c]	13 [c]	Mg^{26} $(2)\rightarrow(1)$

[a] Ki 51. [b] Ki 53a. [c] Br 56e. [d] Ca 56. [e] See also Ki 54a.

Resonance at $E_n=85$ kev has $J=3/2^-$ by measurements of total cross section of enriched Mg^{24} (Ta 56) and by measurements of the differential scattering cross section at 90° and 180° (Bl 56). Resonances at $E_n=275$ and 430 kev also assigned to Mg^{24}; it is proposed that they have $J=1/2^-$ and $3/2^-$, respectively (Fi 51).

IV. $Mg^{24}(d,p)Mg^{25}$ $Q_m=5.107$

Ground-state $Q=5.097\pm0.007$ Mev (En 52b). Ten levels in Mg^{25} (see Table X) found from magnetic analysis with targets of natural magnesium. Assignment of proton groups based on comparison with Mg^{25} levels found from $Al^{27}(d,\alpha)Mg^{25}$ and with proton groups from (d,p) reactions on targets enriched in Mg^{25} and Mg^{26} (En 52a). Also En 54a. Above 4.6 Mev, four incompletely resolved levels found from Al absorption measurements (Ho 53), and several more from low-resolution magnetic analysis (Kr 53a); Table X.

Measurements of γ rays in coincidence with protons which have passed through an Al absorber of variable thickness indicate transitions from the first four excited states of Mg^{25} to ground state and from second and fourth excited states to first one (Mu 56). See Reaction VII for conclusions about spins and parities from γ branching ratios.

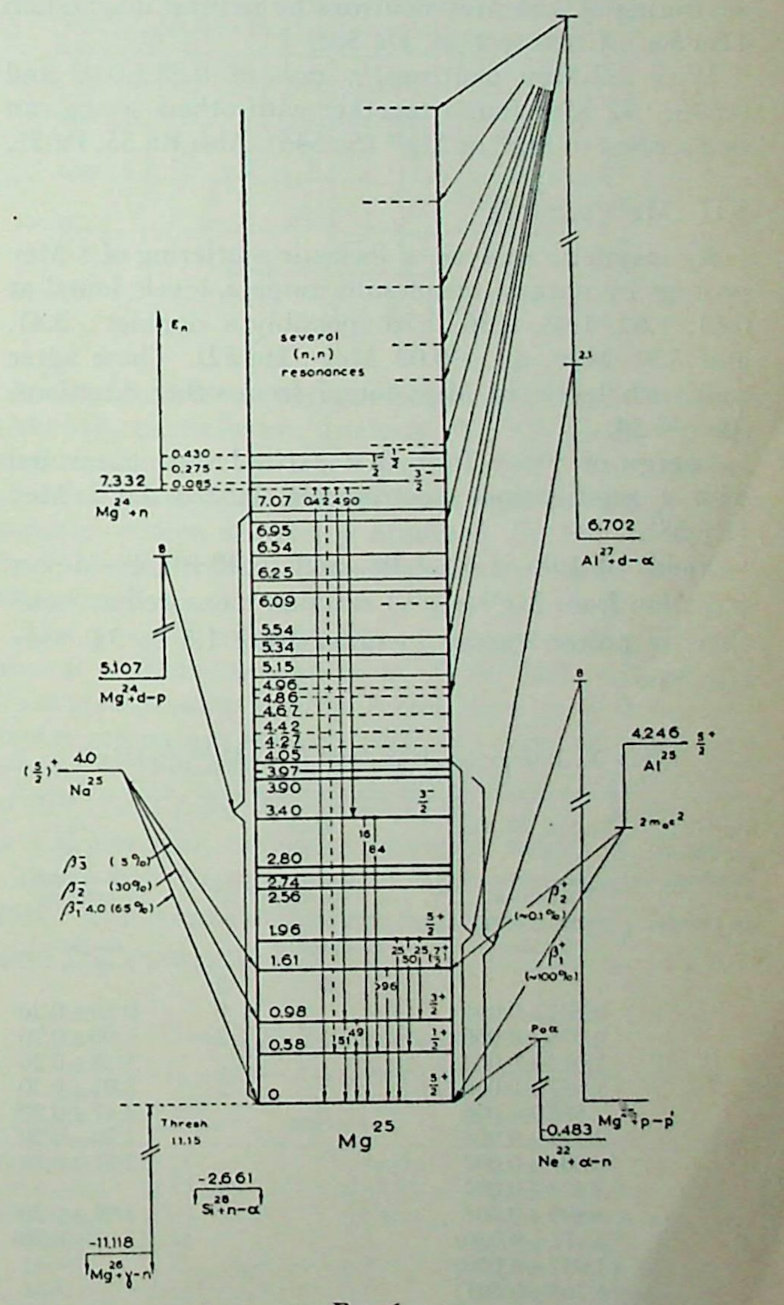


FIG. 6.

Angular distribution measurements yield l_n values in Table X. Angular distribution of protons associated with the 1.61-Mev level is isotropic. Relative intensities of proton groups and neutron capture probabilities, Ho 53.

Angular distribution of protons from (d,p) reaction to 3.40-Mev level and angular correlation between protons and gammas from this level measured at several deuteron energies between 2 and 4 Mev. Results in

accordance with a $3/2^-$ assignment (Wi 55, also Co 56a).

Transition probabilities in $Mg^{24}(d,p)$ and (d,n) reactions, Fu 54.

V. $Na^{25}(\beta^-)Mg^{25}$ See Na^{25}

VI. $Mg^{25}(n,n')Mg^{25}$

A γ ray of 1.616 ± 0.016 Mev results from inelastic scattering of 2.56 Mev neutrons by natural magnesium (Da 56c). Cross section, Da 56c.

With 3.2-Mev neutrons, γ rays of 0.38 ± 0.02 and 0.59 ± 0.02 Mev found together with others which can be ascribed to Mg^{24} or Mg^{26} (Sc 54d). Also Ra 55, Pa 55.

VII. $Mg^{25}(p,p')Mg^{25}$

By magnetic analysis of inelastic scattering of 8-Mev protons by natural magnesium targets, levels found at 0.61, 1.62, 1.98, 2.56, 2.76 (possibly a doublet), 3.41, and 3.91 Mev, all ±0.02 Mev (Ha 52). These agree well with levels in Mg^{25} found from other reactions. Also Fi 54.

Energy of γ ray from first excited state measured with a scintillation spectrometer, 0.588 ± 0.005 Mev (Ka 55).

Yields of 1.96, 1.61, 0.98, and (0.40+0.58)-Mev γ radiation from $Mg^{25}(p,p'\gamma)$ reaction measured as functions of proton energy in range from 1.5 to 3.0 Mev (Go 56).

TABLE X. Energy levels in Mg^{25} from $Mg^{24}(d,p)Mg^{25}$.

Author:	En 52b, En 53a	Ho 53		Kr 53a
E_d (Mev):	1.8	8		4.04
Method:	Magnetic analysis	Al absorption		Magnetic analysis
Q_0 (Mev):	5.097 ± 0.007			
	E_x (Mev)	E_x (Mev) [a]	l_n	E_x (Mev)
		0	2	
	0.582 ± 0.006		0	0.50 ± 0.20
	0.976 ± 0.006		2	1.00 ± 0.20
	1.612 ± 0.006			1.58 ± 0.20
	1.957 ± 0.006		2	2.02 ± 0.20
	2.565 ± 0.006			(2.47 ± 0.20)
	2.742 ± 0.008			2.70 ± 0.20
	2.806 ± 0.007			(2.87 ± 0.20)
	3.405 ± 0.007		1	
	3.899 ± 0.008			3.92 ± 0.20
	3.972 ± 0.010			(4.03 ± 0.20)
	(4.052 ± 0.010)			
	(4.265 ± 0.007)			
	(4.421 ± 0.010)			
		4.62 ± 0.05		4.62 ± 0.20
				(4.86 ± 0.20)
				(4.96 ± 0.20)
		5.05 ± 0.07		5.15 ± 0.20
				5.34 ± 0.20
		5.47 ± 0.07		5.54 ± 0.20
				6.09 ± 0.20
				6.25 ± 0.20
		6.40 ± 0.05		6.54 ± 0.20
				6.95 ± 0.20
				7.07 ± 0.20

[a] Ground-state $Q=5.097$ Mev assumed.

TABLE XI. Energy levels in Mg^{25} from $Al^{27}(d,\alpha)Mg^{25}$.

Author:	Sc 50	To 52	En 52b, En 53a
E_d in Mev:	11.1	10.8	2.1
Method:	Al absorption	Nuclear emulsions	Magnetic analysis
Q_0 (Mev):	6.58 ± 0.03	[a]	6.694 ± 0.010
Levels in Mg^{25} in Mev	0.57 ± 0.05	0.58 ± 0.02	0.584 ± 0.006
	0.96 ± 0.05	0.93 ± 0.04	0.977 ± 0.010
	1.63 ± 0.04	1.62 ± 0.03	1.610 ± 0.010
	1.97 ± 0.05	2.09 ± 0.05	1.958 ± 0.010
			2.558 ± 0.010
	2.74 ± 0.04	2.74 ± 0.03	2.729 ± 0.010
			2.791 ± 0.015
	3.36 ± 0.04	3.36 ± 0.03	3.404 ± 0.012
			3.896 ± 0.015
		3.96 ± 0.04	3.960 ± 0.015
	4.01 ± 0.05	4.12 ± 0.04	4.057 ± 0.015
	4.81 ± 0.05	4.87 ± 0.03	
	5.48 ± 0.05	5.56 ± 0.03	
	5.95 ± 0.05	5.93 ± 0.03	
		6.98 ± 0.03	
		7.85 ± 0.04	
		8.62 ± 0.05	
		9.06 ± 0.04	
		9.75 ± 0.04	
		10.78 ± 0.04	
		11.89 ± 0.05	

[a] Ground-state $Q=6.694$ Mev assumed.

By comparing γ spectra at different resonances and by measuring γ-γ coincidences, branching ratios of levels at 0.98, 1.61, and 1.96 Mev obtained (Fig. 6) (Go 56, Go 57a). Spins and parities of lower levels agree with these branching ratios, hyperfine-structure measurements (Ra 53), (d,p) angular distributions, log ft values in decay of Na^{25} and Al^{25}, shell model predictions, and classifications of corresponding levels in the mirror nucleus Al^{25}.

VIII. $Al^{25}(\beta^+)Mg^{25}$ See Al^{25}

IX. $Mg^{26}(\gamma,n)Mg^{25}$ $Q_m=-11.118$

Threshold at 11.15 ± 0.20 (Sh 51a). Cross sections, Ka 54, Na 55, Ye 56; theoretical discussion of observed cross sections, Mo 55.

X. $Al^{27}(d,\alpha)Mg^{25}$ $Q_m=6.702$

Table XI shows good agreement between level positions found with different bombarding energies and detection methods by different authors and also with level positions deduced from $Mg^{24}(d,p)Mg^{25}$ studies (Table X). Also En 54a, Be 55.

Angular distributions, En 54a.

XI. $Si^{28}(n,\alpha)Mg^{25}$ $Q_m=-2.661$

Not observed.

Al^{25}

(Figs. 7 and 8)

I. $Al^{25}(\beta^+)Mg^{25}$ $Q_m=4.26$

Half-life is 7.62 ± 0.13 sec (Ch 53, Hu 54b, also Br48). The β^+ end point is 3.24 ± 0.03 Mev, as measured

with a magnetic lens spectrometer (El 55, also Hu 54c). Log $ft=3.5$.

Decay accompanied by weak 1.58 ± 0.03-Mev γ ray (Ma 55, also Ch 56), indicating roughly 1% of β^+ transitions go to 1.61-Mev level of Mg^{25}. Log ft value for this transition 5.3. No 1.98-Mev γ ray with intensity $>0.5\%$ was detected (St 56).

II. (a) $Mg^{24}(p,\gamma)Al^{25}$ $E_b=2.29$
(b) $Mg^{24}(p,p)Mg^{24}$ $E_b=2.29$
(c) $Mg^{24}(p,p'\gamma)Mg^{24}$ $E_b=2.29$ $E_x=1.368$

Several authors (Ta 46, Gr 50, Ca 53, Kl 54, Gr 55c, Hu 53, Hu 54b, Hu 55, Cr 56, Ag 56, Li 56a, Va 56b, Va 56c, Ch 56) have observed resonances for the $Mg^{24}(p,\gamma)$ reaction, either by measuring yield of radioactive Al^{25} (Gr 50, Hu 54b, Hu 55), or by measuring yield of capture γ rays (all others). Some resonances found with targets of natural magnesium (Ta 46, Sm 57) ascribed to $Mg^{25}(p,\gamma)Al^{26}$ reaction (Hu 55). Resonances for elastic scattering of protons by Mg^{24} observed for proton energies between 0.4 and 3.9 Mev (Mo 51); partial-wave analysis (Ko 52) yields width and classification of eight levels in Al^{25}. For proton energies between 2.0 and 3.0 Mev, eight resonances for inelastic proton scattering observed (Li 56a, Le 56a), three correspond to those found from elastic proton scattering.

The γ decay of first three resonances studied by several authors; references in Table XII, column 3. Coincidences and angular correlations in γ decay of higher resonances studied with two large scintillation crystals; most assignments of spins and parities to levels of Al^{25} result from this work (Li 56a, Li 57). Resonant proton energies, excitation energies, modes of decay, widths, and classifications of $Mg^{24}+p$ resonances in Table XII. Available information on branching ratios, spins and parities in Fig. 8.

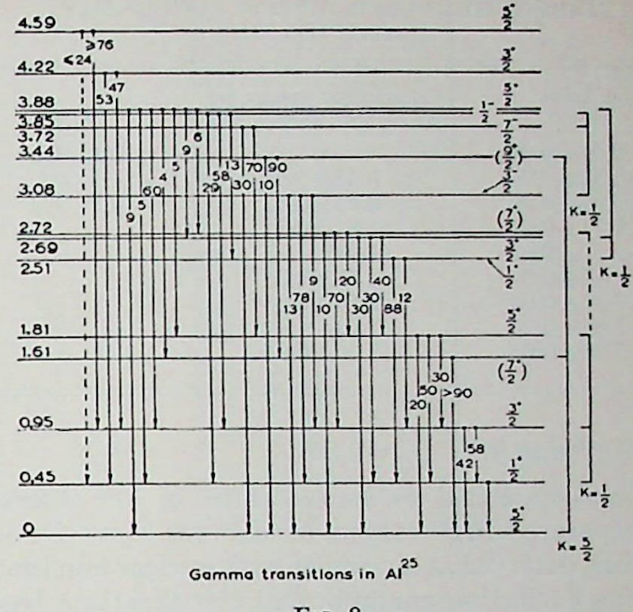


FIG. 8.

TABLE XII. Resonances in $Mg^{24}+p$.

E_p (lab) (kev)	E_x (Al^{25}) (Mev)	Decay [a]			Γ (kev)	J, π
225.5±0.2 [b]	2.51			γ [c,g,h–j]	<1 [m]	1/2+ [i]
418.4±0.5 [b]	2.69			γ [c,d,k]	<1 [e]	3/2+ [c,d,k,l]
824.9±0.4 [c]	3.08	p_0 [e]		γ [c,d,l]	1.5 [n]	3/2− [c,n]
1200 [d]	3.44			γ [d]	<10 [d]	(9/2)+ [d]
1490 [e]	3.72	p_0 [e]		γ [b]	0.3 [n]	7/2− [d]
1620 [e]	3.85	p_0 [e]		γ [b]	36 [n]	1/2− [n]
1660 [e]	3.88	p_0 [e]		γ [d]	0.1 [n]	5/2+ [d]
2010 [e]	4.22	p_0 [e]	p_1 [d,f]	γ [d]	0.15 [n]	3/2+ [d,f]
2400 [e]	4.59	p_0 [e]	p_1 [d,f]	γ [d]	0.3 [n]	5/2+ [d,f]
2720 [d]	4.90		p_1 [d]		<10 [d]	≥5/2 [d]
2890 [d]	5.06		p_1 [d,f]		<10 [d]	1/2, (3/2+) [f]
(2920) [d]	(5.09)		p_1 [d]			
2930 [d]	5.10		p_1 [d,f]		~50 [d]	5/2+ [f]
3140 [e]	5.30	p_0 [e]			200 [n]	3/2− [n]
3660 [e]	5.80	p_0 [e]	p_1 [f]			3/2− [f]
	6.14 [n]		p_1 [f]			1/2, (3/2+) [f]

[a] The symbols p_0, p_1, and γ refer to elastic proton scattering, inelastic scattering to Mg^{24} (1), and proton capture, respectively.
[b] Hu 54b. Hu 55. [c] Ag 56. [d] Li 56a. [e] Mo 51. [f] Le 56a. [g] Ca 53. [h] Cr 56. [i] Va 56b. [j] Ch 56. [k] Va 56c. [l] Gr 55c. [m] Ta 46. [n] Ko 52.

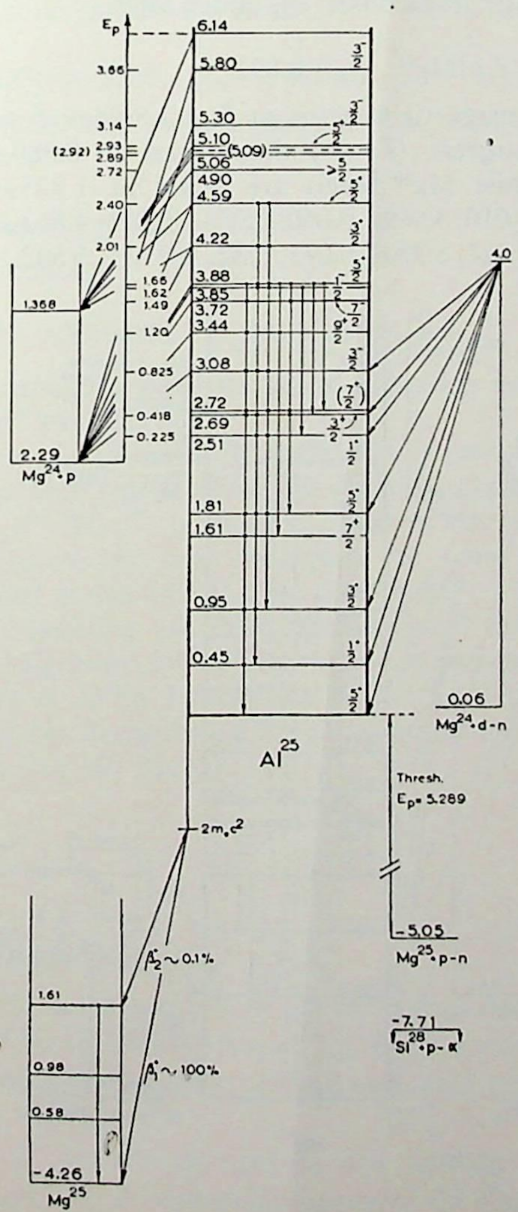


FIG. 7.

TABLE XIII. Levels in Al^{25} from $Mg^{24}(d,n)Mg^{25}$ reaction (Go 53).

Q value (Mev)	Excitation energy in Al^{25} (Mev)	J, π
0.07 ± 0.06	0	$3/2^+, 5/2^+$
-0.38 ± 0.06	0.45 ± 0.03	$1/2^+$
-0.88 ± 0.05	0.95 ± 0.03	$3/2^+, 5/2^+$
-1.74 ± 0.04	1.81 ± 0.04	
-1.87 ± 0.04	(1.94)	
-2.44 ± 0.04	2.51 ± 0.05	
-2.67 ± 0.04	2.70 ± 0.05	
-2.85 ± 0.04	(2.92)	
-3.04 ± 0.03	3.09 ± 0.06	$1/2^-, 3/2^-$

III. $Mg^{24}(d,n)Al^{25}$ $Q_m=0.06$

Ground state and six excited states of Al^{25} observed with separated Mg^{24} targets at $E_d=4.0$ Mev (Go 53). Neutrons detected at six angles with nuclear emulsions. Q values excitation energies, and classifications based on Butler analysis in Table XIII.

IV. $Mg^{25}(p,n)Al^{25}$ $Q_m=-5.05$

Threshold measured as $E_p=5.1$ Mev (Bl 51), 5.25 ±0.1 Mev (Sc 54b), and 5.289 ± 0.025 Mev (Ki 55a). Latter value corresponds to $Q=-5.084\pm0.024$ Mev.

V. $Si^{28}(p,\alpha)Al^{25}$ $Q_m=-7.71$

Not observed.

GENERAL REMARKS

There is a striking correspondence between excitation energies, spins, parities, and γ-branching ratios of the low-lying levels in the mirror nuclei Mg^{25} and Al^{25}. The γ-branching ratios in Al^{25}, about which a large amount of information is available, can be explained on the basis of the collective model if it is assumed that the levels below 4 Mev belong to four rotational bands with $K=5/2, 1/2, 1/2$, and $1/2$ as in Fig. 8. Ground-state M1 transitions from levels of the $K=1/2$ series are attenuated by factors of approximately 20 as compared to M1 transitions between levels in the same bands (Li 56a, Go 56).

Mg^{26}

(Fig. 9)

I. $Ne^{22}(\alpha,\alpha)Ne^{22}$ $E_b=10.635$

Thirteen sharp resonances observed in neon elastic cross section for $E_\alpha=2$–4 Mev at four different angles. Two resonances at $E_\alpha=3.245$ Mev ($\Gamma=2.5\pm0.5$ kev) and $E_\alpha=3.418$ Mev ($\Gamma=3.2\pm0.5$ kev) assigned to Ne^{22}. For both, $J=3^-$ found from partial wave analysis (Go 54c). They correspond to Mg^{26} levels at 13.380 and 13.526 Mev.

II. $Na^{23}(\alpha,p)Mg^{26}$ $Q_m=1.842$

Levels found with α particles from radioactive sources, En 54a.

Yield of 1.83-Mev γ ray de-exciting first Mg^{26} level measured for $E_\alpha=1.9$–3.6 Mev (Te 54; for resonances, Al^{27}). For p-γ coincidence measurements at $E_\alpha=7.8$ Mev yielding γ-decay modes of four lowest Mg^{26} levels, En 54a. Also Br 55b.

III. $Mg^{25}(n,\gamma)Mg^{26}$ $Q_m=11.118$

Cross section, Hu 55c.

Gamma rays from thermal neutron capture in natural magnesium observed with a pair spectrometer (Ki 51, Ca 56) and with a two-crystal scintillation spectrometer (Br 56e). From neutron binding-energies of Mg^{25}, Mg^{26}, and Mg^{27}, all transitions with $E_\gamma>7.4$ Mev must be assigned to Mg^{26}. Several other transitions fit well between known levels of Mg^{26}. Energies, intensities, and proposed assignments in Table IX.

IV. $Mg^{25}(n,n)Mg^{25}$ $E_b=11.118$

Resonances in total cross section of samples enriched in Mg^{25} observed at $E_n=20$ kev ($\Gamma\sim1$ kev), 81 kev ($\Gamma\sim9$ kev), and 101 kev ($\Gamma\sim10$ kev) (Ta 56).

For Mg^{25} levels observed, under Mg^{25}.

V. $Mg^{25}(d,p)Mg^{26}$ $Q_m=8.893$

From magnetic analysis at $E_d=1.8$ Mev (using enriched targets), Q value determined as 8.880 ± 0.012 Mev, while Mg^{26} levels are found at 1.825 ± 0.015, 2.972 ± 0.010, 3.969 ± 0.010, 4.353 ± 0.011, 4.863 ± 0.011, 4.924 ± 0.011, 5.270 ± 0.011, 5.322 ± 0.011, 5.502 ± 0.011,

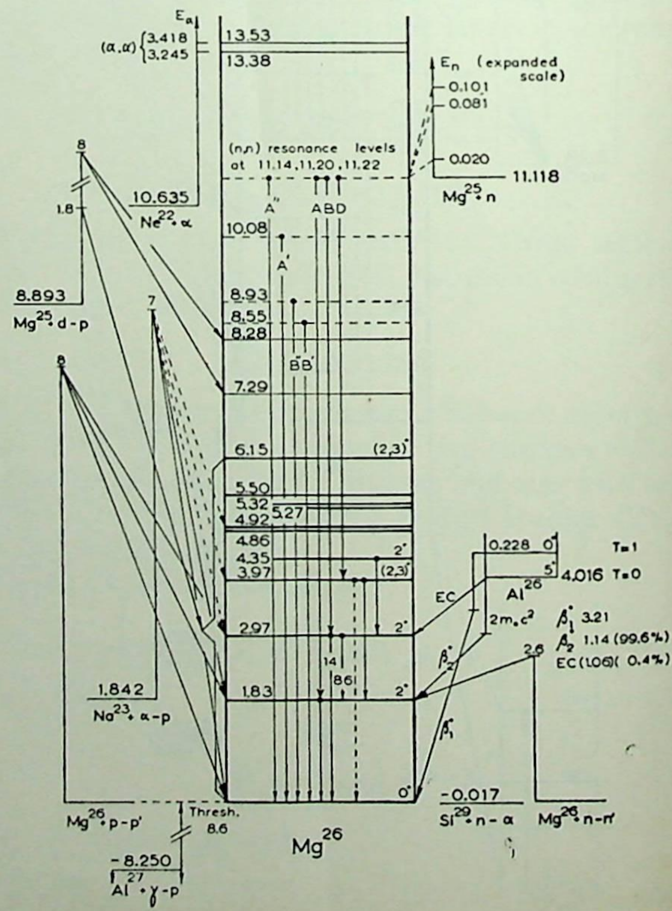


FIG. 9.

and 6.147 ± 0.011 Mev (En 52). Levels from Al absorption or nuclear emulsion methods, En 54a, also Kr 53a.

Angular distribution study at $E_d=8$ Mev indicates neutron orbital angular momentum $l_n=0$ associated with transitions to the 2.97-, 3.97-, 4.35-, and 6.15-Mev levels ($J=2^+$ or 3^+) and both $l_n=0$ and $l_n=2$ with transitions to 1.83-Mev level. Two levels at 7.29 ± 0.06 and 8.28 ± 0.06 Mev found (Ho 53).

Computation of (d,p) reduced widths, Fu 54.

VI. $Mg^{26}(n,n')Mg^{26}$

Inelastic scattering of 2.56-Mev neutrons on natural magnesium gives a 1.820 ± 0.018-Mev γ ray which must be assigned to Mg^{26} (Da 56c). Also Sc 54d, Ra 55, Pa 55. Resonances, under Mg^{27}.

VII. $Mg^{26}(p,p')Mg^{26}$

Magnetic analysis at $E_p=8$ Mev gives levels at 1.83 ± 0.02 and 2.96 ± 0.02 Mev (Ha 52), also En 54a, Fi 54.

VIII. $Al^{26}(\beta^+)Mg^{26}$ See Al^{26}.

IX. $Al^{27}(\gamma,p)Mg^{26}$ $Q_m=-8.250$

Threshold and cross section, En 54a, also Da 56d, Mo 55.

X. $Si^{29}(n,\alpha)Mg^{26}$ $Q_m=-0.017$

Not observed.

GENERAL REMARKS

For jj-coupling computation of excitation energy of first Mg^{26} level, Th 56.

Angular distribution measurements of protons from $Mg^{25}(d,p)Mg^{26}$ reaction limit spin of the 4.35-Mev level to 2^+ or 3^+. The 3^+ possibility, however, can be eliminated because this level belongs to the same isobaric spin triplet as the 2^+ 4.55-Mev level in Al^{26}.

Same reasoning can be applied to Mg^{26} first level. The 2^+ assignment found from unique second-forbidden character of $Al^{26}(\beta^+)$ transition to this level (see Al^{26} Reaction Ia) strengthened by 2^+ assignment to analogous level at 2.07 Mev in Al^{26}.

Al^{26}

(Fig. 10)

I. (a) $Al^{26}(\beta^+)Mg^{26}$ $Q_m=4.016$

Half-life is about 10^6 yr as estimated from $Mg^{25}(d,n)Al^{26}$ yield (Si 54). About 5×10^4 yr would follow from known β^+ end point and an estimated log $ft=13$, (En 54c, Ka 55).

Main β^+ branch proceeds to first level in Mg^{26} at 1.83 Mev. The β^+ end point measured by scintillation spectrometer as $\sim$1 Mev (Si 54), 1.17 ± 0.05 Mev

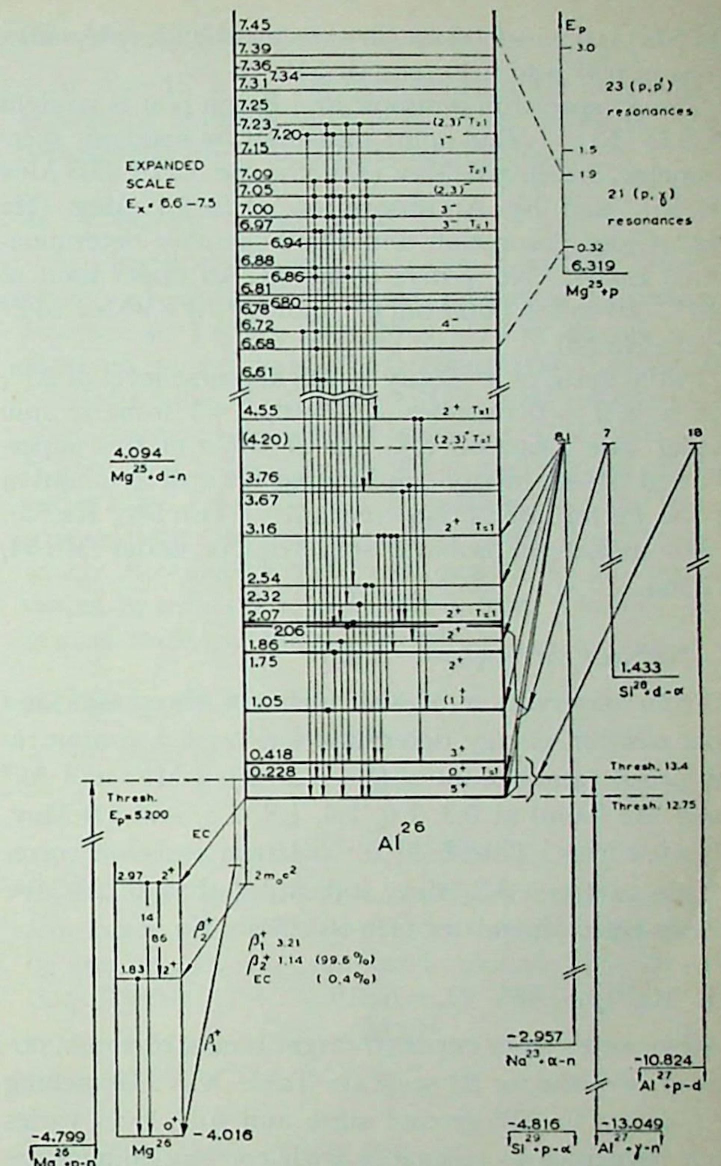


FIG. 10.

(La 55), and 1.14 ± 0.01 Mev (Fe 56), and by Al absorption as 1.30 ± 0.15 Mev (Ha 55b). Shape of β^+ spectrum is unique second forbidden (La 55, Fe 56), which determines spin of first Mg^{26} level (which should be 2^+ or 3^+ from $Mg^{25}(d,p)Mg^{26}$ angular distribution measurements) as 2^+, and that of Al^{26} ground state [a $(d_{5/2}^{-1}, d_{5/2}^{-1})$ state with $J\leq5$ and positive parity] as 5^+. The γ-ray de-exciting first Mg^{26} level measured as 1.9 Mev (Si 54), 1.76 ± 0.1 Mev (La 55), 1.82 Mev (Ha 55b), and 1.84 ± 0.01 Mev (Fe 56). It is in coincidence with β^+ (Ha 55b, Fe 56).

A weak branch proceeds to 2.97-Mev Mg^{26} level, probably by electron capture. A 2.91-Mev γ ray observed with intensity 0.004 relative to that of 1.82-Mev γ ray (Ha 55b). Also a 1.1-Mev γ ray found in coincidence with 1.82-Mev γ ray (Fe 56). A reported 0.717-Mev γ ray (Ha 55b) which cannot be fitted in the Mg^{26} level scheme is probably instrumental (Fe 56).

(b) $Al^{26m}(\beta^+)Mg^{26}$ $Q_m=4.244$

Half-life averaged from three measurements (6.49 ±0.10 sec Ka 51, 6.68 ± 0.11 sec Hu 54b, 6.5 ± 0.1 sec

Ha 54a) is 6.55±0.06 sec, in agreement with determinations of 6.47 and 6.40 sec (Gr 54).

The β^+ spectrum is simple and Fermi plot is straight (Ka 55, El 55). End point measured by magnetic spectrometer, 3.20±0.05 Mev (Ka 55), and 3.21±0.03 Mev (El 55), and by Al absorption, 3.2±0.1 Mev (Ha 54a). Older absorption and cloud-chamber determinations, En 54a. No γ rays observed. An upper limit of 0.01% given for potential β^+ branch to 1.83-Mev Mg^{26} level (Ma 55).

Initial state of β^+ decay is 228 kev first level of Al^{26}, which is $T_z=0$ member of $J=0^+$ $T=1$ isobaric spin triplet (see Reaction X). The ft value of this super-allowed $0^+ \rightarrow 0^+$ decay can be used for a determination of the Fermi interaction constant g_F (En 54c, Ka 55).

Discussions of isobaric spin triplets under Mo 54, Wi 56a.

II. $Na^{23}(\alpha,n)Al^{26}$ $Q_m=-2.957$

From observation of slow neutron thresholds and from neutron energy determinations with a proton recoil proportional counter, $Q=-2.9\pm0.2$ Mev and Al^{26} levels are found at 0.3, 1.0, 1.4, 1.8, 2.5, and 2.9 Mev, all ±0.2 Mev. Threshold for positron emission corresponds to $Q=-3.2$ Mev, indicating that 0.228-Mev level is positron emitter (Do 56). Also, En 54a.

III. $Mg^{25}(p,\gamma)Al^{26}$ $Q_m=6.319$

Resonances from enriched target bombardments, observed in γ and/or β^+ yield, in Table XIV. Branching of γ decay to Al^{26} ground state and first level varies from resonance to resonance with corresponding large variation of β^+/γ ratio. Absence of β^+ emission at 437- and 958-kev resonances is notable. Proton resonance energies averaged to obtain corresponding Al^{26} excitation energies. Also in Table XIV are resonance strengths $(2J+1)\Gamma_p\Gamma_\gamma/(\Gamma_p+\Gamma_\gamma)$ as observed from thick target γ-ray yields.

Natural magnesium targets, En 54a, Sm 57.

Spectra of γ rays with relative intensities measured at 317-, 392-, 437-, 496-, 571-, and 597-kev resonances (Kl 54 superseding Kl 54a), and at 686-, 724-, 777-, 816-, 934-, and 959-kev resonances (Gr 56e), while observed γ rays (without relative intensities) also given at 724- and 959-kev resonances (Ka 55). Gamma-gamma coincidence measurements (Ka 55, Gr 56e), and γ-ray angular distribution measurements, (Gr 56e).

Strong direct transitions to Al^{26} ground state only observed at 437-kev resonance (Kl 54), while first level at 228 kev is fed directly only at 934-kev resonance (Gr 56e). Although positrons are observed at 1191-kev resonance, decay proceeds by at least 99% through $Mg^{25}(p,p')$ reaction to first level in Mg^{25} (Ka 55). Besides other levels found from the $Si^{28}(d,\alpha)Al^{26}$ reaction (Reaction X), new levels at 2.09, 2.32, 2.54, 3.16, 3.67, 3.76, and 4.55 Mev are necessary to explain observed single and coincidence γ-ray spectra (Gr 56e). Also a 5.16-Mev level reported (not indicated in Fig. 10) (Br 56i). Some γ-ray energies given with an accuracy surpassing that of the level differences found from $Si^{28}(d,\alpha)Al^{26}$ reaction. The (2)→(0) transition is given as 419±3 kev (En 54c) and 416±4 kev (Ka 55), the (3)→(1) transition as 820±4 kev (En 54c) and 833±6 kev (Ka 55), and the (6)→(3) transition as 1002±10 kev (En 54c) and 1022±6 kev (Ka 55).

TABLE XIV. Resonances in $Mg^{25}(p,\gamma)Al^{26}$.

Observed:	γ	β^+	γ and β^+	γ and β^+	γ	E_x (Al^{26})	$(2J+1)\frac{\Gamma_p\Gamma_\gamma}{\Gamma_p+\Gamma_\gamma}$ (ev)		J, π, T [a]
Reference:	Kl 54	Ta 54b	Hu 55	Ka 55	Gr 56e	(Mev)	Kl 54	Gr 56e	
E_p (lab): (kev)	321		316.7±0.7			6.608	0.3		
	395	388	391.5±0.5			6.680	0.6		
	441		436.5±0.4			6.723	0.7		4^-
	501	494	495.6±0.6			6.780	0.8		
	518	510	513.4±0.7			6.797	0.8		
			530.4±0.7			6.813			
	580	563				6.856	1.5		
	607	588				6.877	1.1		
	667	650				6.936	0.3		
	688	683		697	675	6.966	1.6	3.5	3^- $(T=1)$
	all ±15	722		730	720	6.999		2.8	3^- $(T=0)$
		777			777	7.050		1.9	$(2,3)^-$ $(T=0)$
		812			820	7.088		1.2	$\pi=-$ $(T=1)$
		880		890		7.154			
		928		933	940	7.201		1.5	1^- $(T=0)$
				958	960	7.225		3.8	$(2,3)^-$ $(T=1)$
		986		990		7.253			
		1043		1046		7.307			
		1081		1086		7.344			
		1098		1105		7.362			
		1132				7.392			
		1197		1185		7.448			
		all ±10–15							

[a] En 54c for assignment at 437-kev resonance

Prominent modes of decay of Al^{26} levels up to $E_x=4.6$ Mev in Fig. 10. Qualitative remarks concerning γ-ray branching percentages, Kl 54, En 54c, Gr 56e. Most spins in Fig. 10 found from γ-ray angular distribution measurements (Gr 56e), some spins and isobaric spins from comparison with Mg^{26} level scheme. Spin assignment to level (3) (probably 1^+) from intensity considerations. Analog of $T=1$, $J=2^+$ or 3^+ level at 3.97 Mev in Mg^{26} not observed. It should have an excitation energy of about 4.20 Mev.

Negative parity resonance levels are each observed to decay predominantly ($\geq 75\%$) either to $T=0$ or to $T=1$ levels. At 686-kev resonance decay proceeds even for 100% to $T=0$ levels. Explained by assuming that resonance levels have T as a fairly good quantum number and that strong transitions from these states are E1 transitions, satisfying the isotopic spin selection rule for self-conjugate nuclei $\Delta T=\pm 1$. Isobaric spins of resonance levels in Table XIV found this way (Gr 56e).

Angular distributions observed at 934-kev $J=1^-$ resonance can only be explained by assuming a large (16 times larger than computed from barrier penetrability alone) f-wave to p-wave ratio (Gr 56e).

IV. $Mg^{25}(p,n)Al^{25}$ $E_b=6.319$ $Q_m=-5.029$

Resonances, En 54a.

Threshold measurements, under Al^{25}.

V. $Mg^{25}(p,p'\gamma)Mg^{25}$ $E_b=6.319$

Yield of 1.96-Mev, 1.61-Mev, 0.98-Mev, and (0.40-Mev+0.58-Mev) γ rays resulting from inelastic proton scattering measured in $E_p=1.5$–3.0-Mev region showing some 23 resonances (not tabulated) (Go 56). Also Ka 55.

VI. $Mg^{25}(d,n)Al^{26}$ $Q_m=4.094$

Thick-target measurements with nuclear emulsions, En 54a.

VII. $Mg^{26}(p,n)Al^{26}$ $Q_m=-4.799$
$Mg^{26}(p,n)Al^{26m}$ $Q_m=-5.027$

A slow-neutron threshold observed at $E_p=5.200$ ± 0.010 Mev corresponding to $Q=-5.006\pm 0.010$ Mev. No threshold corresponding to Al^{26} ground state found (Ki 55a).

Neutron threshold measured as $E_p=5.25\pm 0.1$ Mev (Sc 54b).

VIII. $Al^{27}(\gamma,n)Al^{26}$ $Q_m=-13.049$

Threshold for neutron emission measured as 12.75 ± 0.20 Mev (Sh 51a), and threshold for production of Al^{26m} activity as 14.0 ± 0.4 Mev (Mc 49) and 13.4 ± 0.2 Mev (Ha 54a).

Excitation curves, En 54a, Ha 54a, Mo 55.

IX. $Al^{27}(p,d)Al^{26}$ $Q_m=-10.824$

At $E_p=18$ Mev, angular distribution of deuterons proceeding to one or more of the three lowest Al^{26} states corresponds to $l_n=2$ (Re 56, also Se 56).

X. $Si^{28}(d,\alpha)Al^{26}$ $Q_m=1.433$

At deuteron energies between 5 and 7 Mev and $\theta=90°$, ground-state Q value measured by magnetic analysis as 1.416 ± 0.008 Mev. Lower energy groups observed with Q values of 0.998 ± 0.008, 0.364 ± 0.008, -0.334 ± 0.008, (-0.430 ± 0.015), and -0.648 ± 0.008 Mev, corresponding to Al^{26} levels at 0.418, 1.052, 1.750, (1.856), and 2.064 (Mev. It is assumed that all observed states have $T=0$ (Br 54). Transition violating isobaric spin selection rule also found to lowest $T=1$ level, for which $E_x=228\pm 5$ kev was measured. Its intensity varies, as a function of E_d and θ, from 2 to 65% of the ground-state transition (Br 56d).

XI. $Si^{29}(p,\alpha)Al^{26}$ $Q_m=-4.816$

Not observed.

Si^{26}

(not illustrated)

A 1.7-sec half-life observed from bombardment of aluminum with 23-Mev protons, which might perhaps be attributed to Si^{26} produced through $Al^{27}(p,2p)$ reaction (Ty 54).

Mg^{27}

(Fig. 11)

I. $Mg^{27}(\beta^-)Al^{27}$ $Q_m=2.594$

Half-life averaged from four determinations (En 54a) is 9.45 ± 0.03 min.

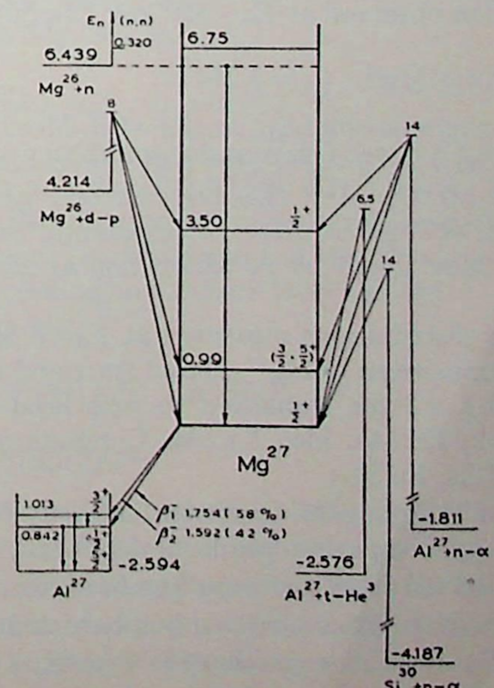


Fig. 11.

TABLE XV. The $Mg^{27}(\beta^-)Al^{27}$ decay.

Reference: Method:	Da 54a Lens spectrometer	Ly 56 Scint. spectrom.	Transition:
E_{β_1} (Mev)	1.754 ±0.011 (58%)		$Mg^{27}\rightarrow(1)$
E_{β_2} (Mev)	1.592 ±0.022 (42%)		$Mg^{27}\rightarrow(2)$
E_{γ_1} (Mev)	0.834 ±0.008	0.834 (70%)	(1)→(0)
E_{γ_2} (Mev)	1.015 ±0.007	1.015 (30%)	(2)→(0)
E_{γ_3} (Mev)	a	0.175 ±0.010 (0.66%)	(2)→(1)

a Intensity of internal conversion electrons $<10^{-3}$ per disintegration.

The β^- decay proceeds to two lowest Al^{27} excited states. Both levels are deexcited mainly by direct γ transitions to Al^{27} ground state, although a weak transition (2)→(1) has also been observed. Table XV gives recent determinations of β^- and γ energies and intensities, En 54a for older work. The two main γ rays are not coincident (Ma 54a, Ly 56); γ_1 is coincident with β_1^-, and γ_2 with β_2^- (Ma 54a). The two β^- transitions are allowed: log $ft=4.75$ and 4.78 for β_1^- and β_2^-, respectively. Computation of matrix elements for these transitions, Fe 55. Also Na 56a.

A reported ground-state transition (Da 53) was later shown to be absent (Da 54a, Ma 54a, Ko 54a).

II. $Mg^{26}(n,\gamma)Mg^{27}$ $Q_m=6.439$

Cross sections, En 54a and Hu 55c. A recent measurement of thermal neutron activation cross section (25±2 mb; Ly 56) appreciably smaller than older values.

Energies of a number of γ rays from capture of thermal neutrons in natural magnesium determined by pair spectrometer (Ki 51, Ki 53a) (Table IX). Line F_1 ($E_\gamma=6.440\pm0.008$ Mev) results from capture in Mg^{26} and represents ground-state transition.

III. $Mg^{26}(n,n)Mg^{26}$ $E_b=6.439$

For samples enriched in Mg^{26}, a resonance in total cross section observed at $E_n\sim320$ kev (Ta 56).

IV. $Mg^{26}(d,p)Mg^{27}$ $Q_m=4.214$

From magnetic analysis at $E_d=1.8$ Mev, Q value measured as 4.207±0.006 Mev and a level observed at $E_x=0.987\pm0.006$ Mev (En 52a). Second group with Q value of 0.71±0.05 Mev corresponding to level at $E_x=3.50$ Mev found by Al absorption at $E_d=8$ Mev (Ho 53).

Angular distributions measured at $E_d=8$ Mev yield $l_n=0$ for transitions to Mg^{27} ground state and to second level, and $l_n=2$ for transition to first level (Ho 53). Older work, En 54a, also Kr 53a. Computation of reduced widths, Fu 54.

Ratio of stripping to compound nucleus formation measured by comparing yields of Mg^{27} and Na^{24} activities observed from deuteron bombardment of magnesium to those from neutron bombardment of aluminum. Ratio has a maximum of 8–9 for $E_d=1$–2 Mev (Ne 56).

V. $Al^{27}(n,p)Mg^{27}$ $Q_m=-1.811$

Proton spectra and angular distributions at $E_n=14$ Mev (Ha 56a, Ov 56, Al 57a, Co 56b). Transitions observed to known Mg^{27} levels, some indication of transitions to 5.7- and 6.9-Mev levels (Ha 56a). Also En 54a.

VI. $Al^{27}(t,He^3)Mg^{27}$ $Q_m=-2.576$

Observed (En 54a).

VII. $Si^{30}(n,\alpha)Mg^{27}$ $Q_m=-4.187$

Cross section, En 54a.

Al^{27}

(Fig. 12)

I. $Na^{23}(\alpha,n)Al^{26}$ $E_b=10.092$

Resonances, En 54a. Q value, under Al^{26}.

II. $Na^{23}(\alpha,p\gamma)Mg^{26}$ $E_b=10.092$

Partially resolved resonances in $E_\alpha=1.8$–3.7-Mev region observed in yield of 1.83-Mev γ ray. They correspond to Al^{27} levels at 11.75, 11.91, 12.07, 12.23, 12.27,

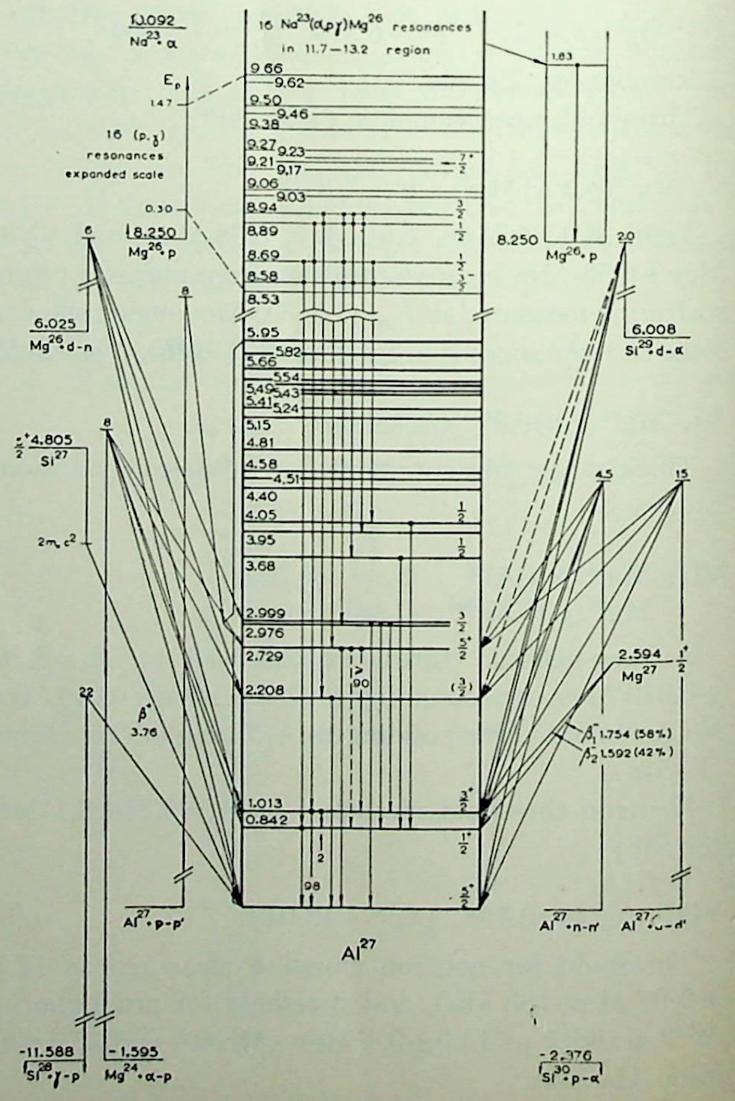


FIG. 12.

TABLE XVI. Resonances in $Mg^{26}(p,\gamma)Al^{27}$.

Reference:	Ta 46	Kl 54	Ru 54a	Hu 55	E_x (Al^{27}) (Mev)	$\frac{\Gamma_p\Gamma_\gamma}{\Gamma_p+\Gamma_\gamma}$ (ev)	J, π
E_p (lab) (kev)	290±3	300±15			8.529		
	336±1.5	340±15	343	338.5±0.5	8.575	0.4 ±0.1 [b]	3/2− [a,b]
	451±2	458±15	450	454.2±0.3	8.687	1.4 ±0.3 [b]	1/2 [b]
			662		8.887	0.7 ±0.2 [b]	1/2 [b]
			720		8.943	0.21±0.05 [b]	3/2 [b]
			813		9.033		
			840		9.059		
			954		9.169		
			992		9.205		7/2+ [c]
			1015		9.227		
			1056		9.267		
			1172		9.379		
			1255		9.459		
			1295		9.497		
			1425		9.622		
			1465		9.661		
			all ±10 kev				

[a] Hu 56.
[b] Va 56d, En 56b.
[c] Al 57.

12.34, 12.46, 12.68, 12.71, 12.78, 12.84, 12.87, 12.99, 13.07, and 13.14 Mev (Te 54).

III. $Mg^{24}(\alpha,p)Al^{27}$ $Q_m=-1.595$

Resonances under Si^{28}. From comparison of resonances in $Mg^{24}(\alpha,p)Al^{27}$ and $Al^{27}(p,\alpha)Mg^{24}$, $Q=-1.613 \pm 0.010$ Mev (Ka 52).

Direct measurement at $E_\alpha=8$ Mev with a proportional counter, $Q=-1.63\pm0.07$ Mev, and Al^{27} levels at 0.85, 1.06, 2.17, and 2.64 Mev, all ±0.04 Mev (Gr 56d). Also, Br 55b.

IV. (a) $Mg^{25}(d,n)Al^{26}$ $E_b=17.143$
(b) $Mg^{25}(d,p)Mg^{26}$ $E_b=17.143$

Possible resonance, En 54a.

V. $Mg^{26}(p,\gamma)Al^{27}$ $Q_m=8.250$

Resonances in γ-ray yield from enriched target bombardments (Kl 54, Ru 54a, Hu 55) in Table XVI, in which also those resonances observed by Tangen (Ta 46) from natural magnesium bombardments which can be assigned to Mg^{26}. Corresponding Al^{27} excitation energies computed from weighted average of resonant proton energies. Thirty-five resonances reported in the $E_p=0.80$–2.36-Mev region (Al 57).

Gamma-ray spectra, γ-ray relative intensities, angular distributions, and γ-γ coincidences measured at 338, 454-, 662-, and 720-kev resonances (Va 56d, En 56b), and at some unspecified resonance (Al 57). Resonance spins found from this work in Table XVI.

Parity of 338-kev resonance obtained from polarization measurement of 7.7-Mev γ ray de-exciting the resonance to level (1), which determined character of γ ray as E1 (Hu 56).

Spins of a number of lower levels

0.842 Mev $J=1/2$, 2.976 and/or 2.999 Mev $J=3/2$,
2.208 Mev $J=(3/2)$, 3.68 Mev $J=1/2$,
2.729 Mev $J=5/2$, 4.05 Mev $J=1/2$.

Energies of γ rays de-exciting first and second Al^{27} levels: 0.840±0.006 Mev and 1.012±0.008 Mev, Va 56d, En 56b.

Angular distribution measurements (Al 57) yield $J=3/2$ for 1.013-Mev level, and, very probably, $J=1/2$ for 0.842-Mev level. The 2.729-Mev level decays for more than 90% to 1.013-Mev level. Angular correlations measured at 986-kev resonance in agreement with $J=7/2^+$ or $3/2^+$ for the resonance level and $5/2^+$ for 2.729-Mev level. The $7/2^+$ assignment to the resonance is favored because no direct transitions to 0.842- and 1.013-Mev levels observed (intensity at most 3% of ground-state transition).

Also En 54a, Sm 54.

VI. $Mg^{26}(d,n)Al^{27}$ $Q_m=6.025$

Nuclear emulsion work at $E_d=1.0$ and 6.0 Mev gives Al^{27} levels at 1.0, 2.3, 2.9, 6.8, 8.2, 9.5, and 10.0 Mev. From angular distribution measurements $l_p=0$ assigned to transitions proceeding to 6.8-Mev level and $l_p=1$ to those proceeding to 8.2-Mev level (Tr 56).

Also En 54a.

VII. $Mg^{27}(\beta^-)Al^{27}$ See Mg^{27}.

VIII. $Al^{27}(n,n')Al^{27}$

Numerous authors have published γ-ray energies and intensities from inelastic neutron scattering measured with a scintillation spectrometer, Table XVII. Existence of 1.20-Mev γ ray reported at $E_n=3.2$ Mev seems doubtful because not found in careful work at $E_n=2.56$ Mev. The 1.91-Mev γ ray observed at $E_n=3.7$ Mev also

TABLE XVII. Gamma rays from $Al^{27}(n,n')Al^{27}$.

Reference: E_n (Mev): Transition	Da 56c 2.56 $E\gamma$ (Mev)	σ (mb)	Kl 54 2.7 $E\gamma$ (Mev)	Sc 54d 3.2 $E\gamma$ (Mev)	Ro 55a 3.7 $E\gamma$ (Mev)	Gr 55d 4.5 $E\gamma$ (Mev)
(2)→(1)	0.166±0.003	1.6± 0.4		[a]		
(1)→(0)	0.840±0.008	64 ± 7	0.850±0.025	0.89±0.03	0.835±0.01	0.84±0.02
(2)→(0)	1.017±0.010	142 ±13	1.025±0.03	1.05±0.03	1.02 ±0.01	1.02±0.03
(3)→(2)?				1.20±0.05[b]		
(4)→(2)				1.70±0.08	1.72 ±0.02	
(4)→(1)?					1.91 ±0.02	
(3)→(0)	2.21 ±0.02	87 ± 8	2.23 ±0.04	2.20±0.18	2.22 ±0.03	2.27±0.08

[a] The 0.05±0.02-Mev γ ray which is also reported might perhaps be identified with 31 kev (1)→(0) transition in Al^{28} following neutron capture.
[b] Might be pair peak of 2.21-Mev γ ray.

doubtful because it would correspond to a $5/2^+ \rightarrow 1/2^+$ E2 transition.

Thresholds for 0.84-, 1.01-, and 2.21-Mev γ rays are equal, within experimental limits, to γ-ray energies, proving these are ground-state transitions (Ki 54). Also En 54a, Gr 55b, Gr 55d, Pa 55, Ra 55, Ra 55a, Ta 55, Be 56, Cr 56a, Mo 56c, Sm 56, We 56a.

Angular distributions of elastically scattered neutrons, Li 55a, Be 56, Cu 56, Do 56a, Hi 56a.

IX. $Al^{27}(p,p')Al^{27}$

Levels found by magnetic or electrostatic analysis in Table XVIII. Excitation energies reported in Re 52 consistently higher by an amount slowly increasing up to 75 kev for the highest levels than the values found from high-resolution work given in Br 54b. Groups corresponding to levels reported in Re 52 at 4.996 and 5.107 Mev are probably from carbon contamination on front and back side of target (Br 54b). Also En 54a, Ma 54b.

TABLE XVIII. Levels in Al^{27} from $Al^{27}(p,p')Al^{27}$.

Reference: E_p (Mev): Angle: Method:	Sh 51 2–4 164° magn. analysis	Re 52 8.0 90° magn. analysis	Do 53 2.3 135° electrostat. analysis	Br 54b 5.6–8.4 90° magn. analysis	Va 57d 2–4 90° magn. analysis
E_x (Mev)	0.82	0.844	0.843±0.002	0.842	0.842±0.003
	1.04	1.016		1.013	1.013±0.003
	2.23	2.259		2.213	2.205±0.004
	2.75	2.782		2.732	2.727±0.004
				2.977	2.975±0.004
		3.046		3.001	2.998±0.004
		3.736		3.677	
		4.018		3.954	
		4.115		4.054	
		4.473		4.403	
		4.575		4.505	
		4.647		4.576	
		4.875		4.807	
		4.996			
		5.107			
		5.220		5.150	
		5.341		5.242	
				5.410	
		5.501		5.425	
		5.565		5.491	
		5.620		5.544	
		5.736		5.659	
		all ±0.020		5.821	
				5.951	
				all ±0.006	

Yield of 0.84- and 1.01-Mev γ rays from inelastic proton scattering measured in $E_p=1.40$–2.77-Mev region (Pa 54, Br 57a). Resonances under Si^{28}. A $P_2(\cos\theta)$ term observed in angular distribution of 1.01-Mev γ ray, showing that spin of 1.01-Mev level is not 1/2. This level is de-excited for 2.2% by a 0.175-Mev cascade transition. Angular correlation of 0.175- and 0.84-Mev γ rays shows no significant deviation from isotropy pointing to a spin $J=1/2$ for 0.84-Mev level (Br 57a). An upper limit of 10^{-10} sec is found for half-life of these γ transitions from recoil techniques (Sw 56).

Angular distributions of protons scattered elastically (En 54a, Da 54, Hi 55, Sh 56, Da 56a, Ge 56a), and inelastically (no groups resolved (Gu 54)).

X. $Al^{27}(d,d')Al^{27}$

Low-resolution analysis of inelastically scattered deuterons, En 54a.

Angular distributions for transitions to 2.21- and 2.73-Mev levels at $E_d=15$ Mev (Ha 56c), and for transitions to 2.21 and 2.73 levels and to doublet at 2.99 Mev at $E_d=9$ Mev (Hi 56b). Angular momentum transfer connected with these three transitions determined as $l_d=2$ fixing parity of corresponding levels as positive (Hi 56b).

Elastic scattering angular distribution at $E_d=14$ Mev, Go 50.

XI. $Al^{27}(\alpha,\alpha')Al^{27}$

Elastic scattering angular distributions at $E_\alpha=19$ and 40 Mev, Bl 55a, Ig 56, Bl 56a. Inelastic scattering to first two Al^{27} levels (unresolved) also observed (Bl 56a).

XII. $Si^{27}(\beta^+)Al^{27}$ See Si^{27}

XIII. $Si^{28}(\gamma,p)Al^{27}$ $Q_m=-11.588$

Cross section, Jo 55, Mo 55a.

No activity with a half-life in the μsec or msec range found from bombardment of natural silicon with 22-Mev bremsstrahlung (Ve 56a).

XIV. $Si^{29}(d,\alpha)Al^{27}$ $Q_m=6.008$

Magnetic analysis of α particles observed at 90° at $E_d=1.8$ and 2.0 Mev from targets enriched in Si^{29} content yields ground-state Q value of 5.994 ± 0.011 Mev and levels in Al^{27} at 0.837 ± 0.016 and 1.007 ± 0.013 Mev, and possibly at 2.21 ± 0.03 and 2.74 ± 0.02 Mev (Va 52).

XV. $Si^{30}(p,\alpha)Al^{27}$ $Q_m=-2.376$

Not observed.

Si^{27}

(not illustrated)

I. $Si^{27}(\beta^+)Al^{27}$ $Q_m=4.805$

Three measurements of half-life with errors of at most 0.1 sec in rather bad agreement (4.92 ± 0.1 sec El 41; 4.45 ± 0.05 sec Su 53; 4.05 ± 0.10 sec Hu 54). Average value is 4.46 ± 0.18 sec.

The β^+ end point measured by cloud chamber as 3.74 Mev (Mc 40) and 3.54 ± 0.1 Mev (Ba 40), and by scintillation spectrometer as 3.48 ± 0.10 Mev (Bo 51), and 3.76 ± 0.08 Mev (Hu 54). No γ rays reported. Decay is apparently superallowed ($\log ft=3.6$) determining spin and parity of Si^{27} as $5/2^+$.

II. $Mg^{24}(\alpha,n)Si^{27}$ $Q_m=-7.183$

Observed, En 54a.

III. $Al^{27}(p,n)Si^{27}$ $Q_m=-5.588$

Recent measurements of threshold are $E_p=5.792$ ±0.010 Mev (Ki 55a, superseding Ki 53) and $E_p=5.798$ 0.008 Mev (Ma 55a).

Angular distribution measured at $E_p=23$ Mev (Co 55b).

IV. $Si^{28}(\gamma,n)Si^{27}$ $Q_m=-17.176$

Threshold determined as 16.8 ± 0.4 Mev (Mc 49) and 16.9 ± 0.2 Mev (Su 53). Cross section, En 54a, Ka 54, Mo 55a. No activity in the μsec or msec range observed from bombardment of natural silicon with 22-Mev bremsstrahlung (Ve 56a).

V. $Si^{28}(p,d)Si^{27}$ $Q_m=-14.951$

Cross section, Se 56.

Mg^{28}

(not illustrated)

I. $Mg^{28}(\beta^-)Al^{28}$ $Q_m=1.810$

Mg^{28} has been produced through $Mg^{26}(\alpha,2p)$, Si^{30} $(\gamma,2p)$, $Si^{30}(p,3p)$, and other spallation reactions, and through $Mg^{26}(t,p)$ reaction.

Half-life, averaged from six measurements (En 54a, Sh 54d), in good mutual agreement, 21.43 ± 0.11 hr.

TABLE XIX. Gamma rays from $Mg^{28}(\beta^-)Al^{28}$.

Reference:	Sh 54d		Wa 53		Iw 53
	$E\gamma$ (Mev)	Rel. int.	$E\gamma$ (Mev)	Rel. int.	$E\gamma$ (Mev)
γ_1	0.0319 ± 0.001	0.96	0.0322	($\sim$0.70)	
γ_2	0.400 ± 0.01	0.31			0.391 ± 0.005
γ_3	0.949 ± 0.01	0.29			0.95
γ_4	1.346 ± 0.01	0.70			1.35

The β^- spectrum is simple and has allowed shape (Ma 53a, Ol 54). End point 0.40 ± 0.06 Mev (Sh 54d), 0.3 Mev (Jo 53), 0.39 ± 0.05 Mev (Wa 53) (all measured by Al absorption), 0.418 ± 0.010 Mev (Ma 53a), and 0.459 ± 0.002 Mev (Ol 54), latter two measured by magnetic spectrometer. $\log ft=4.4$.

Four γ rays observed by scintillation spectrometer. Energy and intensity (in photons per disintegration) in Table XIX. Gamma rays γ_2, γ_3, and γ_4 coincident with γ_1, and γ_2 with γ_3. No measurable delay between γ_1 and γ_4 ($\tau<2\times10^{-9}$ sec) (Sh 54d). Accepting more accurate γ-ray intensities in Sh 54d, conversion coefficient of γ_1 can be computed from intensities as $\alpha_K=0.032\pm0.066$, which determines the character of γ_1 as M1.

From γ-ray energy and coincidence measurements, β^- decay proceeds to 1.37-Mev level in Al^{28}, which decays through 0.97- and 0.03-Mev levels. By the isobaric spin selection rule, excluding $\Delta T=\pm1$, $0^+\rightarrow0^+$, β^- transitions, spin and parity of 1.37 Mev level determined as 1^+. Assuming $J=3^+$ for Al^{28} ground state and $J=2^+$ for 0.03-Mev level then leaves possibilities $J=0^+$ or 2^+ for 0.97-Mev level to explain absence of β^--decay to this level and observed γ-ray branching from 1.37-Mev level.

Al^{28}

(Fig. 13)

I. $Al^{28}(\beta^-)Si^{28}$ $Q_m=4.648$

Half-life, averaged from three most accurate measurements (En 54a, Co 56), in good mutual agreement, is 2.28 ± 0.02 min.

The β^- spectrum is simple with allowed shape. Two most accurate magnetic spectrometer determinations of end point are 2.865 ± 0.010 Mev (Mo 52) and 2.878 ±0.014 Mev (Ol 54). $\log ft=4.9$.

Each β^- particle followed by one γ quantum. Best determinations of γ-ray energy are 1.782 ± 0.010 Mev (magn. spectrometer, Mo 52), and 1.769 ± 0.010 Mev (scint. spectrometer, Sh 54d). Also En 54a, Ma 54a, Na 54a, Na 56a.

II. $Mg^{25}(\alpha,p)Al^{28}$ $Q_m=-1.204$

From enriched-target measurement at $E_\alpha=8$ Mev, with proportional counter for proton detection, Q value is -1.29 Mev. Levels in Al^{28} reported at 1.00, 1.57, 2.18, 2.54, and 2.96 Mev, all ±0.04 Mev (Gr 56d). Also En 54a.

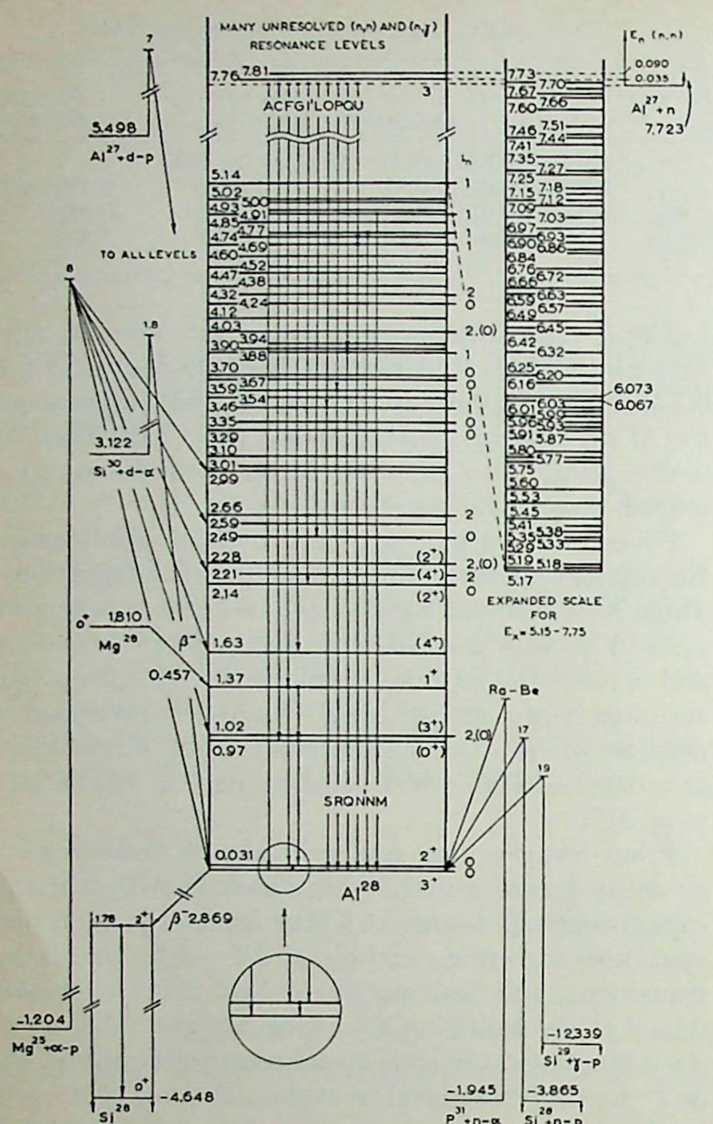


FIG. 13.

III. $Al^{27}(n,\gamma)Al^{28}$ $Q_m=7.723$

Cross section, Hu 55c.

Twenty-nine γ rays resulting from capture of thermal neutrons in aluminum found by pair spectrometer (Ki 51, Ki 53d), Table XX. Strong γ-ray A represents ground-state transition. Its great intensity is a notable exception to rule that only E1 ground-state transitions are intense (Ki 52b). Impossible to fit other γ rays uniquely into the complicated Al^{28} level scheme. Lines C, F, G, I', J, L, and all lower energy lines could be transitions from capturing state to known Al^{28} levels. Lines B, B', D, E, F', G', H, K, and L cannot be explained this way but have to be transitions from level to ground or from level to level.

From $Al^{27}(d,p)Al^{28}$ reaction (see Table XXI) the parity of levels (17), (19), (22), (33), (35), (37), and (41) is negative. It is possible to explain many strong γ rays (e.g., M, N, N', O, P, Q, R, S, U, V) of Table XX as E1 transitions to and from these levels (En 56d).

With two-crystal scintillation spectrometer, γ rays from thermal neutron capture observed with the following energies and relative intensities (in photons per 100 captures): $E_\gamma=0.97\pm0.03$ Mev (10), 2.26 ± 0.03 Mev (14), ~2.5 Mev (11), ~2.8 Mev (14), and ~3.1 Mev (17). Last three γ rays not resolved (Br 56e).

Also En 54a.

IV. $Al^{27}(n,n)Al^{27}$ $E_b=7.723$

Cross section, Hu 55c. Resonances at $E_n=35\pm2$ kev ($\Gamma=1.2\pm0.5$ kev), and 90 ± 4 kev ($\Gamma=7\pm2$ kev) have both $J=3^+$ (Hu 55c, To 55a, Bl 56). In Cr 57, however, $J=2^+$ is assigned to 35-kev resonance.

Angular distributions, Hu 55c, Hi 56a, Na 57.

V. $Al^{27}(d,p)Al^{28}$ $Q_m=5.498$

Fifty proton groups corresponding to Al^{28} levels up to $E_x=6.3$ found by high-resolution magnetic analysis ($\theta=90°$) at deuteron energies up to 2.1 Mev (En 51, En 52a, En 53a). Later work with still better resolution, at $E_d=6$ and 7 Mev and at several angles of observation, 100 groups found in a region of excitation in Al^{28} up to neutron threshold (Bu 56). Ground-state Q value is 5.494 ± 0.008 Mev (En 52a, En 53a), and 5.502 ± 0.010 Mev (Bu 56).

Excitation energies of Al^{28} levels reported in Bu 56 in Table XXI. Corresponding Q values all have errors of 10 kev. Excitation energies might have errors ranging from 5 kev for lower to 10 kev for highest levels. Groups corresponding to levels above threshold for neutron emission of low intensity and not tabulated. Their positions correspond approximately to apparently unresolved resonances in Al^{27} total neutron cross section (Hu 55c).

Angular distributions covering the region from 5° to 60° measured at $E_d=6$ Mev for transitions leading to levels (0) through (41) (En 56a, En 56). About half have stripping character, while the others are fairly isotropic. The l_n values of stripping distributions and corresponding reduced widths in relative measure (times $(2J+1)$) in Table XXI. For transitions to level (3), where $l_n=2$ and $l_n=0$ contributions are mixed, the $l_n=0$ contribution is 12%. For levels (8) and (25) it

TABLE XX. Gamma rays from thermal neutron capture in aluminum (Ki 51, Ki 53d).

γ ray	Energy (Mev)	Intensity in photons per 100 captures	γ ray	Energy (Mev)	Intensity in photons per 100 captures
A	7.724 ± 0.006	35	L	5.21 ± 0.02	3
B	7.34 ± 0.04	0.7	M	4.94 ± 0.05	1.4
B'	6.98 ± 0.04	0.7	N	4.79 ± 0.02	9
C	6.77 ± 0.02	1.4	N'	4.66 ± 0.05	7
D	6.61 ± 0.03	0.4	N''	4.45 ± 0.02	3
E	6.50 ± 0.03	0.4	O	4.29 ± 0.02	8
F	6.33 ± 0.02	2	P	4.16 ± 0.02	6
F'	6.22 ± 0.03	0.7	P'	4.06 ± 0.04	3
G	6.13 ± 0.02	2	Q	3.88 ± 0.02	8
G'	6.01 ± 0.05	0.7	R	3.62 ± 0.02	6
H	5.89 ± 0.04	1.0	S	3.46 ± 0.02	4
I	5.78 ± 0.03	1.4	T	3.29 ± 0.02	5
I'	5.60 ± 0.02	3	U	3.02 ± 0.05	15
J	5.41 ± 0.03	3	V	2.84 ± 0.03	13
K	5.32 ± 0.03	1.0			

is very small. Possible $l_n=3$ admixtures to $l_n=1$ distributions undetected because of limited angular region of observation.

From angular distribution measurements it is reasonably certain that ground-state doublet has the $(d_{5/2}, s_{1/2})$ configuration predicted by shell model. For a 3^+ ground state and a 2^+ first level the intensity ratio of transitions to these states should be 1.4 (assuming equal reduced widths) while experiment (at 6.0 Mev) yields 1.95 (see Table XXI) (En 53a, En 54, En 56).

TABLE XXI. Levels in Al^{28} from $Al^{27}(d,p)Al^{28}$ (Bu 56, En 56a, En 56).[a]

Level number	E_x (Mev)	l_n	$(2J+1)$ × reduced width [b]	Level number	E_x (Mev)
(0)	0	0	700	(50)	5.445
(1)	0.0312±0.0005 [c]	0	360	(51)	5.525
(2)	0.973	...	...	(52)	5.596
(3)	1.017	2, (0)	600	(53)	5.746
(4)	1.372	...	...	(54)	5.766
(5)	1.633	...	...	(55)	5.802
(6)	2.143	0	220	(56)	5.867
(7)	2.207	2	160	(57)	5.909
(8)	2.279	2, (0)	400	(58)	5.931
(9)	2.490	0	40	(59)	5.960
(10)	2.589	...	...	(60)	5.989
(11)	2.663	2	400	(61)	6.012
(12)	2.988	...	...	(62)	6.027
(13)	3.011	...	...	(63)	6.067
(14)	3.102	...	...	(64)	6.073
(15)	3.294	0	20	(65)	6.163
(16)	3.347	0	25	(66)	6.201
(17)	3.461	1	320	(67)	6.247
(18)	3.537	...	...	(68)	6.322
(19)	3.591	1	410	(69)	6.424
(20)	3.669	0	12	(70)	6.446
(21)	3.704	0	35	(71)	6.485
(22)	3.878	1	120	(72)	6.569
(23)	3.900	...	...	(73)	6.591
(24)	3.935	...	...	(74)	6.626
(25)	4.030	2, (0)	300	(75)	6.657
(26)	4.115	...	...	(76)	6.719
(27)	4.243	0	20	(77)	6.760
(28)	4.315	2	110	(78)	6.835
(29)	4.383	...	...	(79)	6.856
(30)	4.466	...	...	(80)	6.896
(31)	4.518	...	...	(81)	6.934
(32)	4.595	...	...	(82)	6.970
(33)	4.685	1	650	(83)	7.025
(34)	4.741	...	...	(84)	7.090
(35)	4.767	1	500	(85)	7.121
(36)	4.845	...	...	(86)	7.149
(37)	4.906	1	330	(87)	7.180
(38)	4.928	...	...	(88)	7.247
(39)	4.999	...	...	(89)	7.274
(40)	5.019	...	...	(90)	7.345
(41)	5.138	1	340	(91)	7.408
(42)	5.168			(92)	7.444
(43)	5.179			(93)	7.460
(44)	5.191			(94)	7.505
(45)	5.289			(95)	7.596
(46)	5.331			(96)	7.655
(47)	5.346			(97)	7.669
(48)	5.377			(98)	7.700
(49)	5.405			(99)	7.731

[a] *Note added in proof.*—Improved and extended angular distribution data in H. A. Enge, M.I.T., Laboratory for Nuclear Science Progress Report, November 30, 1956.
[b] Reduced width given in relative measure.
[c] Other errors, see text.

TABLE XXII. Rotational bands in Al^{28} (Sh 56a).

E_x (Mev)	0	0.031	0.97	1.02	1.37	1.63	2.14	2.21	2.28
J, π	3^+	2^+	0^+	3^+	1^+	4^+	2^+	2^+	4^+
K	3	2	0	2	1	3	1	0	2

A 31.4±1.0-kev γ ray observed following $Al^{27}(d,p)Al^{28}$ reaction at $E_d=0.7$ Mev both by proportional counter and by scintillation spectrometer (Sm 51a). Half-life of first level measured with a recoil method, yielding $(2.1\pm0.4)\times10^{-9}$ sec (Se 56a).

Also En 54a, Kr 54b, Fu 54, Be 55.

VI. $Mg^{28}(\beta^-)Al^{28}$ See Mg^{28}

VII. $Si^{28}(n,p)Al^{28}$ $Q_m=-3.865$

Cross section, En 54a, Co 56, Ma 56.

VIII. $Si^{29}(\gamma,p)Al^{28}$ $Q_m=-12.339$

Cross section, Ka 54.

IX. $Si^{30}(d,\alpha)Al^{28}$ $Q_m=3.122$

Ground-state $Q=3.120\pm0.010$ Mev measured with enriched targets and magnetic analysis at $E_d=1.8$ Mev (St 51). An α-particle group leading to 31-kev level in Al^{28} observed at several deuteron energies up to $E_d=2.0$ Mev, but with insufficient resolution to determine corresponding excitation energy (En 53a).

X. $P^{31}(n,\alpha)Al^{28}$ $Q_m=-1.945$

Cross section, En 54a.

GENERAL REMARKS

Arguments for a rotational character of nine lowest states in Al^{28}, Sh 56a. There are four bands (Table XXII); those with $K=2$ and $K=3$ have $(d_{5/2}, s_{1/2})$ configuration, while those with $K=0$ and $K=1$ belong to $(s_{1/2}, s_{1/2})$. The 4^+ spin assigned in Sh 56a to 2.28-Mev level conflicts with 2^+ or 3^+ assignment from $Al^{27}(d,p)Al^{28}$ angular distribution measurements. It would be better to assign 4^+ $(K=2)$ to 2.21-Mev level and 2^+ $(K=0)$ to 2.28-Mev level.

Si^{28}

(Fig. 14)

I. $Mg^{24}(\alpha,p)Al^{27}$ $E_b=9.990$ $Q_m=-1.595$

In experiments with thin separated $Mg^{24}F_2$ targets and magnetic analysis of protons, resonances observed at $E_\alpha=3.214$, 3.420, 3.448, 3.502, 3.660, 3.737, and 3.827 Mev corresponding to levels in Si^{28} at 12.745, 12.921, 12.945, 12.992, 13.127, 13.193, and 13.270 Mev. Same levels found in inverse reaction $Al^{27}(p,\alpha)Mg^{24}$, when a Q value of 1.613±0.010 Mev is assumed. Intensities of corresponding peaks in good agreement, as

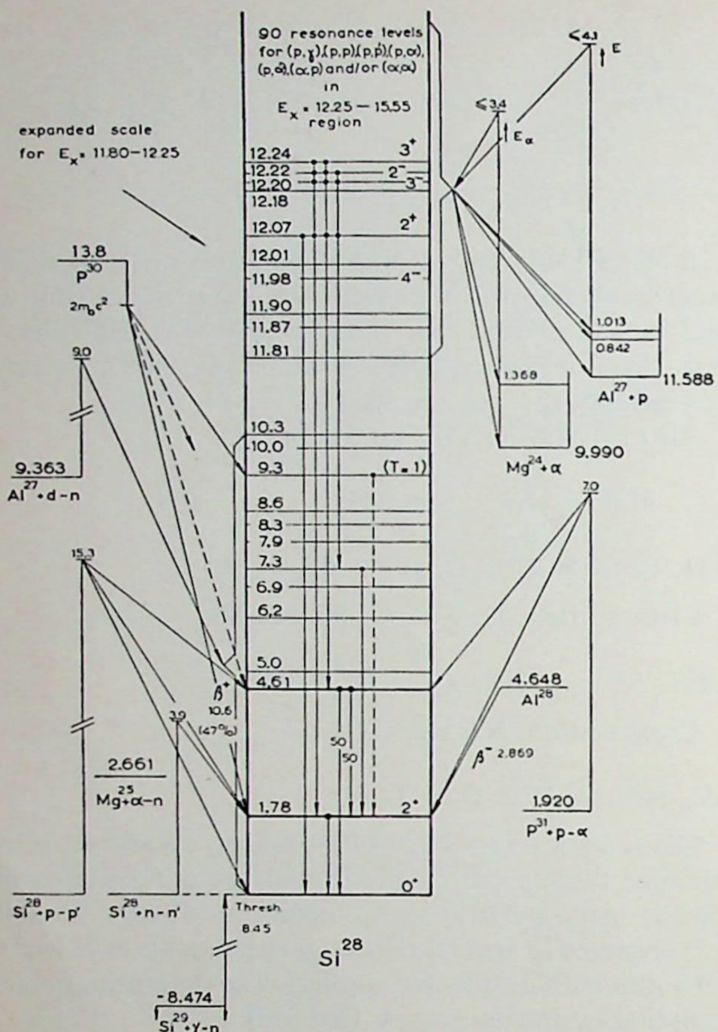


FIG. 14.

suggested by principle of detailed balancing. A weak resonance found at $E_\alpha=2.84$ Mev without corresponding resonance in (p,α) reaction (Ka 52). Also Al^{27}, for Q-value measurements.

II. $Mg^{24}(\alpha,\alpha)Mg^{24}$ $E_b=9.990$

Resonances in yield of elastically scattered α particles observed, again in good agreement with many resonances in the $Al^{27}(p,\alpha)Mg^{24}$ and $Al^{27}(p,p)Al^{27}$ reactions, (Ka 52). Partial-wave analysis yields following spins and parities for compound nucleus: $J=0^+$ for resonances at $E_\alpha=2.84$ and 3.00 Mev, and $J=2^+$ for resonance at $E_\alpha=2.93$ Mev. S-wave α-scattering resonance at $E_\alpha=3.07$ Mev corresponds to S-wave proton scattering resonance at $E_p=1.45$ Mev. No single J value of compound nucleus can explain this behavior. No (α,p) nor (p,α) transitions observed at $E_\alpha=3.07$ Mev. Anomaly near $E_\alpha=3.25$ Mev probably is caused by two overlapping levels in Si^{28} (Ka 52).

III. $Mg^{25}(\alpha,n)Si^{28}$ $Q_m=2.661$

En 54a.

IV. $Al^{27}(p,\gamma)Si^{28}$ $Q_m=11.588$

Resonances up to $E_p=1.38$ Mev in Table XXIII. Energies of six lowest resonances in good agreement with those in Ta 46. Thirty-one γ-ray resonances in $E_p=0.46$–2.6-Mev region in Pl 40. Fourteen are below $E_p=1.4$ Mev, their energies in fair agreement with energies in Br 47a. Resonances at 504-, 630-, and 728-kev decay also by α emission (Ru 53).

Spin and parity assignments (Ru 54) result from measurements of γ-ray spectra, γ-ray angular distributions, and α-particle yield. The 1.78-Mev level in Si^{28} has $J=2^+$. At most resonances, transitions observed to 7.1- and 4.6-Mev levels, which both de-excite by cascades through 1.78-Mev level (Ru 54). The 4.6-Mev level, however, also deexcites to the Si^{28} ground state with intensity about equal to that of cascade (Go 57).† Parity of 652-kev resonance has been determined from a γ-ray polarization measurement (Hu 56).

Precision measurements of energy of prominent 991-kev resonance yield 993.1±1 kev (He 49) and 990.8 ±0.2 kev (Bu 56b). Width is 100 ev (Be 49) and 50±30 ev (Bu 56b). From Doppler-shift measurements of 1.78-Mev γ ray, half-life of 1.78-Mev level follows as either

TABLE XXIII. Resonances in $Al^{27}(p,\gamma)Si^{28}$ below $E_p=1.38$ Mev.

E_p (kev)	E_x (Si^{28}) (Mev)	Relative intensity [d,e]	J, π
226.3±1.5 [a]	11.806	0.005	
294.1±0.5 [a]	11.871	0.015	
325.6±0.4 [a]	11.902	0.080	
404.7±0.4 [a]	11.978	0.30	4^- [b]
438.5±0.5 [a]	12.011	0.050	
504.0±0.6 [a]	12.074	2.0	2^+ [b]
609 [c]	12.175	0.4	
630 [c]	12.195	8.2	3^- [b]
652 [c]	12.217	3.3	2^- [b,f]
677 [c]	12.241	1.3	3^+ [b]
728 [c]	12.290	3.0	
733 [c]	12.295	4.2	
738 [c]	12.300	0.7	
757 [c]	12.318	3.8	
764 [c]	12.325	4.5	
771 [c]	12.331	11.5	
880 [c]	12.436	0.5	
918 [c]	12.473	4.1	
932 [c]	12.487	3.9	
991 [g,h]	12.544	47.0	
994 [c]	12.547	2.0	
1018 [c]	12.570	7.2	
1083 [c]	12.632	1.5	
1091 [c]	12.640	0.8	
1112 [c]	12.660	13.5	
1165 [c]	12.711	2.4	
1176 [c]	12.722	6.5	
1192 [c]	12.737	...	
1205 [c]	12.750	11.0	
1255 [c]	12.798	13.0	
1268 [c]	12.811	1.0	0^+ [e]
1309 [c]	12.850	14.0	
1320 [c]	12.861	10.5	
1355 [c]	12.895	15.0	
1372 [c]	12.911	105.0	
1379 [c]	12.918	105.0	

[a] Hu 53. [b] Ru 54. [c] Br 47a (errors in proton energies are 0.2%, widths of all resonances observed are ≤1 kev). [d] Ta 46. [e] Ka 52 (from $Mg^{24}(\alpha,\alpha)Mg^{24}$ partial wave analysis). [f] Hu 56. [g] He 49. [h] Bu 56b.

† *Note added in proof.*—Recent measurements, however, show that no such crossover exists in decay of 4.6-Mev level (Go 57b).

$<3\times10^{-14}$ sec or $>9\times10^{-13}$ sec (De 56a). For γ-ray spectrum at this resonance, Ha 55.

Resonances above $E_p=1.4$ Mev in Pl 40, Sh 51, Go 54, Pa 54. Yields of hard and weak γ rays, of elastically and inelastically scattered protons, and of α particles leading to ground state and first excited state in Mg^{24} in Sh 51 for region $E_p=0.6$–4.1 Mev. From published yield curves, resonance energies for 64 resonances in 1.4–4.1-Mev region computed in Al 50, quoted in En 54a. No hard γ rays from capture observed above $E_p=1.8$ Mev. Comparison of resonance strengths (Go 53a) makes it probable that resonances at $E_p=1355$, 1372, and 1379 kev from Br 47a correspond to resonances at 1370, 1385, and 1393 kev in Sh 51, Al 50. Not possible at present to decide which energy scale is better. Angular distributions of γ-ray transitions leading to level (1) measured at 17 resonances in $E_p=0.65$–2.20-Mev region, and of ground-state transition at 771-kev resonance (Go 57b).

Twenty-three resonances, yielding 0.84-Mev and/or 1.01-Mev γ rays from inelastic scattering observed in the $E_p=1.40$–2.77 Mev region (Br 57a).

V. $Al^{27}(p,n)Si^{27}$ $E_b=11.588$ $Q_m=-5.610$

Broad resonances in Si^{27} yield observed at $E_p=6.17$ and 6.37 Mev (Bl 51). See Si^{27} for threshold measurements.

VI. (a) $Al^{27}(p,p)Al^{27}$ $E_b=11.588$
(b) $Al^{27}(p,\alpha)Mg^{24}$ $E_b=11.588$ $Q_m=1.595$

Resonances, Sh 51, Pa 54, quoted under Reaction IV.

VII. $Al^{27}(d,n)Si^{28}$ $Q_m=9.363$

Determinations of excitation energies of levels in Si^{28}, and angular distribution measurements, Table XXIV. Relative intensities of neutron groups, Ca 55. Comparison of $Al^{27}(d,p)$ and $Al^{27}(d,n)$ angular distributions makes it probable that 9.16-Mev level in Si^{28} is lowest $T=1$ state (Ca 55).

With a magnetic pair spectrometer, γ rays resulting from deuteron bombardment ($E_d=4.6$ Mev) of an aluminum target have these energies: $E_\gamma=6.9\pm0.1$, 7.41 ± 0.05, 7.58 ± 0.05, 7.94 ± 0.03, 8.31 ± 0.03, 8.78 ± 0.03, 9.11 ± 0.03, 9.49 ± 0.07, 9.91 ± 0.07, and 10.8 ± 0.2 Mev. Most of these probably follow $Al^{27}(d,n)Si^{28}$ reaction, as levels in Al^{28} above 7.72 Mev and levels in Mg^{25} above 7.33 Mev, which might be reached by (d,p) or (d,α) reaction, are unstable for neutron emission (Be 55).

VIII. $Al^{28}(\beta^-)Si^{28}$ See Al^{28}

IX. $Si^{28}(e,e')Si^{28}$

Differential cross section for elastic scattering of 187-Mev electrons and for inelastic scattering leading to level (1), (He 56).

TABLE XXIV. Neutron groups from the reaction $Al^{27}(d,n)Si^{28}$.

Reference:	Pe 49	Ru 56	Ca 55
Method of detection:	nucl. emuls.	nucl. emuls.	triple ion. chamber
E_d (Mev):	3.68	2.17	9.02
Q_0 (Mev)	9.08 ± 0.20	...	$l_n=2$
E_x (Si^{28}): in Mev	1.78 ± 0.13	1.78 ± 0.10	$l_n=0$
	4.47 ± 0.10	4.54 ± 0.2	} ...
	(4.91 ± 0.21)	4.95 ± 0.2	
	6.11 ± 0.10	6.24 ± 0.06	} $l_n=0$
	6.65 ± 0.14	6.88 ± 0.06	
	(7.10 ± 0.12)	7.39 ± 0.06	...
	(7.55 ± 0.12)	7.89 ± 0.06	$l_n=0$
	8.18 ± 0.10	8.31 ± 0.10	$l_n=0$
		8.57 ± 0.08	
	9.16 ± 0.17	9.37 ± 0.04	$l_n=0$
		10.00 ± 0.10	
		10.25 ± 0.06	$l_n=0$

X. $Si^{28}(n,n'\gamma)Si^{28}$

Gamma rays of 1.29, 1.78 ± 0.02, and 2.19 Mev observed from inelastic neutron scattering ($E_n=0.35$–3.9 Mev) on natural silicon (Da 56). They have to be assigned to deexcitation of first levels in Si^{29}, Si^{28}, and Si^{30}, respectively. A 1.78 ± 0.02-Mev γ ray, Ro 55a.

XI. $Si^{28}(p,p')Si^{28}$

From magnetic analysis at $E_p=5.6$ and 8.4 Mev, first Si^{28} level observed at 1.777 ± 0.010 Mev (Br 54b). At $E_p=15.3$ Mev, a level at 4.6 ± 0.3 Mev found by low-resolution magnetic deflection, Fu 48.

At $E_p=12$ Mev, angular distribution of proton group leading to level (1), Co 55.

Resonances in yield of 1.78 Mev γ ray, P^{29}.

XII. $P^{28}(\beta^+)Si^{28}$ See P^{28}

XIII. $Si^{29}(\gamma,n)Si^{30}$ $Q_m=-8.474$

Threshold is 8.45 ± 0.20 Mev (Sh 51a). Cross section, Ka 54.

XIV. $P^{31}(p,\alpha)Si^{28}$ $Q_m=1.920$

Ground-state Q by magnetic analysis $=1.85\pm0.02$ Mev (Fr 51), 1.909 ± 0.010 Mev (Va 52a), 1.911 ± 0.005 Mev (Va 56), and 1.909 ± 0.010 Mev (En 57b). Levels in Si^{28} observed at 1.771 ± 0.010 and 4.617 ± 0.010 Mev (En 57b). Resonances, S^{32}.

P^{28}

(not illustrated)

I. $P^{28}(\beta^+)Si^{28}$ $Q_m=13.8$

Half-life, averaged from two measurements (Gl 55, Br 54a) in good mutual agreement, is 0.285 ± 0.007 sec.

The β^+ decay quite complicated. Highest energy branch proceeds to 1.78-Mev Si^{28} level with end point of 10.6 ± 0.4 Mev, intensity $47\pm15\%$. Observed gamma

TABLE XXV. Gamma rays in decay of P^{28}.

Reference:	Gl 55 $E\gamma$ (Mev)	Rel. int.	Br 54a $E\gamma$ (Mev)
	1.79 ± 0.02	0.75	1.78 ± 0.04
	$(2.6\ \pm0.2)$		2.67 ± 0.08
			(3.01 ± 0.07)
			(4.26 ± 0.12)
	4.44 ± 0.05	0.10	4.63 ± 0.10
	(4.93 ± 0.08)		4.89 ± 0.09
			(5.16 ± 0.12)
			(5.46 ± 0.10)
	6.14 ± 0.10	0.10	
	6.70 ± 0.12	0.10	6.65 ± 0.11
	7.04 ± 0.08	0.10	7.10 ± 0.12
	7.59 ± 0.15	0.05	(7.44 ± 0.14)
			(7.73 ± 0.14)
			(8.12 ± 0.21)

rays in Table XXV. No delayed α particles observed (Gl 55).

Most of observed high-energy γ rays cannot be fitted uniquely into Si^{28} level scheme. A natural explanation of the 7.59-Mev γ ray, Gl 55, assumes part of the β^+ decay proceeds to 9.3-Mev level which is de-excited through 1.78-Mev level. Superallowed character of such a β^+ transition ($\log ft=3.4$–3.7) would be in accordance with assignment of $T=1$ to 9.3-Mev level (Bo 55, Wi 56a).

Log $ft=4.9$ for main β^+ transition to 1.78-Mev level, equal to that of Al^{28} β^- decay to same level. This is another proof of charge symmetry of nuclear forces (Gl 55).

II. $Si^{28}(p,n)P^{28}$ $Q_m=-14.6$

Threshold for P^{28} yield measured as $E_p=15.6\pm0.3$ Mev (Gl 55) and 15.4 ± 0.5 Mev (Br 54a).

Al^{29}

(not illustrated)

I. $Al^{29}(\beta^-)Si^{29}$ $Q_m=3.98$

Measurements of half-life yield an average of 6.52 ± 0.05 min (En 54a).

The β^- decay proceeds to first and third excited states of Si^{29} (at $E_x=1.28$ and 2.43 Mev) which are de-excited by γ transitions to ground state. No 2.03-Mev γ ray of intensity $\geq 4\%$ observed; no evidence found for cascade de-excitation of 2.43-Mev level through that at 1.28 Mev (Ro 55, Br 56). Energies and relative intensities of β^- and γ rays Table XXVI. Also Br 57.

TABLE XXVI. The $Al^{29}(\beta^-)Si^{29}$ decay.

Author	Method	E_{β_1} (Mev)	E_{β_2} (Mev)	E_{γ_1} (Mev)	E_{γ_2} (Mev)
Me 37	Al abs	2.75			
Be 39	cl. ch.	2.5			
Se 49	Al abs	2.5 (70%)	1.4 (30%)	1.25 ± 0.02	2.35 ± 0.5
Na 54b	scint. sp		1.55 ± 0.1	1.31 ± 0.05	2.42 ± 0.05
Ro 55	scint. sp			1.28 (85%)	2.43 (15%)
Br 56	scint. sp			1.28 (94%)	2.43 (6%)

Log ft values for transitions to 1.28-, 2.03-, and 2.43-Mev levels are <5.2, >6.5, and 5.2, respectively (Br 57).

Also Na 54a.

II. $Mg^{26}(\alpha,p)Al^{29}$ $Q_m=-3.19$

With 8-Mev cyclotron α particles and detection of protons with a proportional counter, ground-state Q value is -2.90 Mev; a probable level at 1.69 ± 0.04 Mev (Gr 56d). Also En 54a, Br 56b.

III. $Al^{27}(t,p)Al^{29}$ $Q_m=8.38$

Observed (En 54a).

IV. $Si^{29}(n,p)Al^{29}$ $Q_m=-3.20$

Cross section, En 54a.

V. $Si^{30}(\gamma,p)Al^{29}$ $Q_m=-13.81$

Cross section, Ka 54, En 54a.

Si^{29}

(Fig. 15)

I. $Mg^{26}(\alpha,n)Si^{29}$ $Q_m=0.017$

En 54a, Na 53a, Br 55b.

II. $Mg^{25}(\alpha,p)Al^{28}$ $E_b=11.135$ $Q_m=-1.204$

En 54a, Na 53a.

III. $Si^{28}(n,\gamma)Si^{29}$ $Q_m=8.474$

Thermal-neutron capture cross section of natural silicon is 130 ± 30 mb; isotopic cross sections of Si^{28}, Si^{29}, and Si^{30} are 80 ± 30, 270 ± 90, and 400 ± 400 mb, and their abundances 92.27, 4.68, and 3.05%, respectively, so approximately 73% of the thermal-neutron captures in natural silicon should occur in Si^{28} (Hu 55c).

Energies and intensities of γ rays from capture of thermal neutrons in natural silicon in Table XXVII. Assignment of most γ transitions based on comparison of γ energy with binding energies and excitation energies in Si^{29}, Si^{30}, and Si^{31} as measured by magnetic analysis of (d,p) reactions. The γ spectrum of the $Si^{28}(n,\gamma)Si^{29}$ reaction dominated by two cascades via levels at 4.93 and 6.38 Mev; see General Remarks.

IV. $Si^{28}(n,n)Si^{28}$ $E_b=8.474$
$Si^{28}(n,n')Si^{28}$ $E_b=8.474$

Total neutron cross section of natural silicon, Hu 55c. Resonances at $E_n=195$ and 570 kev are assigned to Si^{28} and have $J=1/2^+$ and $3/2^-$, respectively (Fi 51, Hu 55c). See Da 56 for inelastic-scattering resonances.

V. $Si^{28}(d,p)Si^{29}$ $Q_m=6.249$

Thirteen proton groups found by magnetic analysis at $\theta=90°$ and E_p up to 2.1 Mev (En 51a, Va 52). Q

TABLE XXVII. Gamma rays from thermal neutron capture in natural silicon.

Reference: Method: Line	Ki 51, Ki 53c Pair spectrometer $E\gamma$ (Mev)	Intensity [a]	Ad 56a Compton el. spectr. $E\gamma$ (Mev)	Intensity [a]	Br 56e Scint. spectr. $E\gamma$ (Mev)	Intensity [a]	Final nucleus and probable transition
A	10.599±0.011	(0.4)	10.59 ±0.03	(0.2)			Si^{30} (*c*)→(0)
B	8.467±0.008	(2[b])	8.482±0.015	(1.6)			Si^{29} (*c*)→(0)
C	7.79 ±0.05	(1)					
D	7.36 ±0.08	(2)	(7.38 ±0.03)	(0.5)			(Si^{30})
E	7.18 ±0.03	(8[b])	7.22 ±0.03	(6.1)			Si^{29} (*c*)→(1)
E'	6.88 ±0.03	(0.7)					
F	6.76 ±0.04	(4)	6.758±0.020	(1.4)			(Si^{30})
G	6.40 ±0.03	(11[b])	6.354±0.015	(9.2)			Si^{29} (12)→(0)
H	6.11 ±0.05	(4)	6.04 ±0.05	(1)			(Si^{29})
I'	5.70 ±0.04	(2)					
I	5.52 ±0.05	(2)					
I''			5.24 ±0.03	(0.8)			
J	5.11 ±0.04	(8[b])	5.118±0.015	(2.3)			(Si^{29})
K	4.933±0.005	(75[b])	4.930±0.010	(37.4)			Si^{29} (9)→(0)
K'	4.60 ±0.08	(4)					
L	4.20 ±0.03	(19)	(4.30 ±0.05)	(2)			
L'			3.976±0.020	(4.2)			(Si^{29})
L''			3.8	(3)			(Si^{30})
L'''			3.667±0.020	(3.2)			
M	3.540±0.006	(60[b])	3.547±0.010	(36.5)	3.6 ±0.06	(49)	Si^{29} (*c*)→(9)
N	2.69 ±0.05	(65)			2.65±0.03	(11)	
O			2.10 ±0.01	(12.8)	2.13±0.03	(19)	Si^{29} (*c*)→(12)
P			1.95 ±0.02	(3.4)			
Q			1.7	(3)			(Si^{29})
R			1.5	(3)			(Si^{29})
S			1.28 ±0.01	(16)	1.26±0.03	(14)	Si^{29} (1)→(0)

[a] Intensity in photons per 100 captures in natural silicon.
[b] Intensity in photons per 100 captures in Si^{28}; see Ki 54a.

values and excitation energies in Table XXVIII agree with values obtained by absorption measurements at $E_d=3.7$ Mev (Mo 50). Excellent agreement between levels found here and from $P^{31}(d,\alpha)Si^{29}$ (Table XXIX).

Angular distributions of most prominent proton groups measured at $E_d=8.2$ Mev; angular momenta of capture neutrons, from stripping theory, in Table XXVIII; these l_n values and neutron capture probabilities discussed on basis of shell model (Ho 53a, Ho 53d). Assignment of $l_n=0$ to ground state was confirmed by measurements at $E_d=9$ Mev (Ca 56a). Polarization of protons leading to p states at 4.93 and 6.38 Mev measured (Ju 56).

TABLE XXVIII. Energy levels in Si^{29} from $Si^{28}(d,p)Si^{29}$.

Author:	Va 52	Ho 53d
E_d (Mev):	1.8–2.1	8.2
Method:	magnetic analysis	Al absorption
Q_0 (Mev):	6.246±0.010	
	E_x (Mev)	l_n
	0	0
	1.278±0.007	2
	2.027±0.007	2
	2.426±0.007	isotropic
	3.070±0.007	2
	3.623±0.007	3
	4.078±0.008	
	4.840±0.008	
	4.897±0.008	
	4.934±0.008	1
	5.946±0.009	
	6.105±0.009	
	6.380±0.009	1

Measurements and discussion of reduced widths for $Si^{28}(d,p)Si^{29}$ and $Si^{28}(d,n)P^{29}$ ground-state transitions, Fu 54, Ca 56a.

Levels at 1.28, 2.03, 2.43, 3.07, 4.93, and 6.38 Mev all decay by γ transitions directly to ground state (Al 49b, Th 54, Ro 55).

Angular correlation of protons and gammas corresponding to 1.28-Mev level differs from that predicted by simple stripping theory (Al 56a).

VI. $Al^{29}(\beta^-)Si^{29}$ See Al^{29}

VII. $Si^{29}(n,n'\gamma)Si^{29}$

The 1.28-Mev level is excited by inelastic scattering of neutrons with energies up to 3.9 Mev (Da 56).

VIII. $Si^{29}(p,p'\gamma)Si^{29}$

Angular distributions of 2.03-Mev γ ray measured at $E_p=2.798$ and 2.934 Mev. Presence of P_4 terms indi-

TABLE XXIX. $P^{31}(d,\alpha)Si^{29}$ reaction (En 51a, Va 52a).

Group	Q value (Mev)	Si^{29} level (Mev)
A	8.158±0.011	0
B	6.885±0.020	1.274±0.010
C	6.126±0.020	2.032±0.014
D	5.727±0.020	2.431±0.015
E	5.086±0.020	3.072±0.016
F	4.539±0.020	3.619±0.017
G	4.080±0.020	4.078±0.018
H	3.221±0.020	4.937±0.020

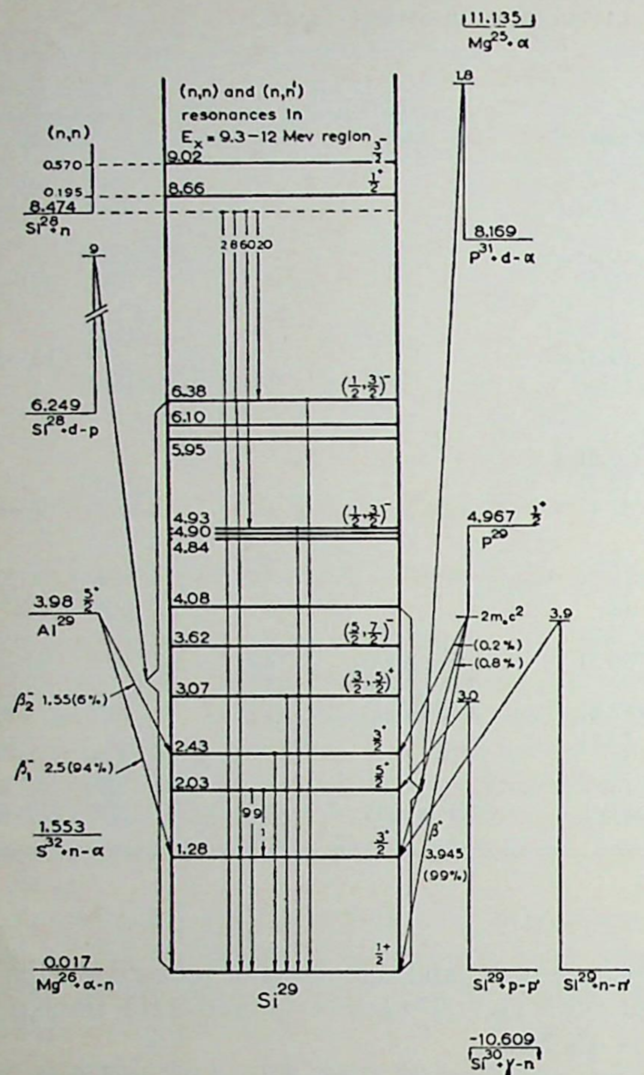


FIG. 15.

cates that 2.03-Mev level has $J=5/2^+$. From coincidence measurements, this level decays for about 0.5% via a cascade through 1.28-Mev level. Angular distributions of 1.28-Mev γ ray at $E_p=2.798$ and 2.922 Mev are consistent with assignment of $3/2^+$ to 1.28-Mev level (Br 56, Br 57).

IX. $P^{29}(\beta^+)Si^{29}$ See P^{29}

X. $Si^{30}(\gamma,n)Si^{29}$ $Q_m=-10.609$

Cross sections, Ka 54.

XI. $P^{31}(d,\alpha)Si^{29}$ $Q_m=8.169$

Eight α-particle groups (Table XXIX) found by high-resolution magnetic analysis at $\theta=90°$ and $E_d=1.8$ Mev (En 51a). Ground-state Q value given as 8.170 ±0.020 Mev later remeasured as 8.158 ± 0.011 Mev (Va 52a). Excitation energies of Si^{29} levels in Table XXIX corrected accordingly. Also Be 55.

XII. $S^{32}(n,\alpha)Si^{29}$ $Q_m=1.553$

En 54a; Hu 55a.

GENERAL REMARKS

The p states at 4.93 and 6.38 Mev are strongly excited, both in $Si^{28}(d,p)Si^{29}$ and in $Si^{28}(n,\gamma)Si^{29}$ reactions; these may be single-particle $p_{3/2}$ and $p_{1/2}$ states predicted by shell model. Ground state and 1.28- and 3.62-Mev levels may be regarded as $s_{1/2}$, $d_{3/2}$, and $f_{7/2}$ states (Ho 53d, Ki 54a).

Spin of ground state of Si^{29} is $1/2^+$ (Ra 53, Wh(53, Wi 54, Og 54). Assignment of $3/2^+$, $5/2^+$, and $3/2^+$ to the levels at 1.28, 2.03, and 2.43 Mev, respectively, of $1/2^+$ to P^{29} (0) and $5/2^+$ to Al^{29} (0) in agreement with observed log ft values, (d,p) and $(p,p'\gamma)$ angular distributions, and γ branching ratios.

Application of strong-coupling collective model to low levels of Si^{29}, Br 57c.

P^{29}

(Fig. 16)

I. $P^{29}(\beta^+)Si^{29}$ $Q_m=4.967$

Half-life is 4.6 ± 0.2 sec (Wh 41) and 4.45 ± 0.05 sec (Ro 55).

The β^+ end point by magnetic spectrometer is 3.945 ±0.005 Mev, yielding $\log ft=3.723\pm0.007$ (Ro 55). Gamma rays of 1.28 and 2.43 Mev in coincidence with positrons observed; branching ratios, log ft values, and proposed classifications in Table XXX. Spins and parities of Si^{29} levels, see Si^{29}, general remarks. Theo-

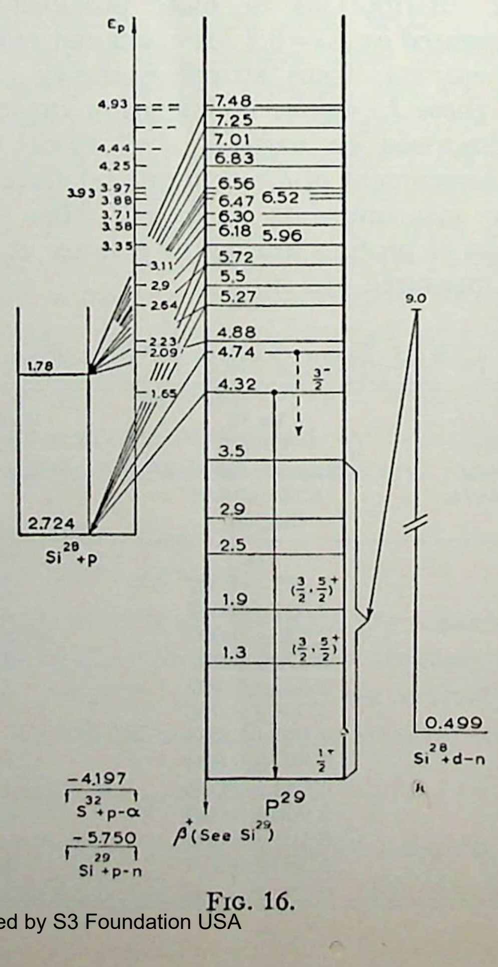


FIG. 16.

TABLE XXX. The $P^{29}(\beta^+)Si^{29}$ decay (Ro 55).

Si^{29} level (Mev)	Branching ratio %	Log ft
2.43	$0.24^{+0.26}_{-0.08}$	4.5
2.03	$\leqslant 0.15$	$\geqslant 5.2$
1.28	0.80 ± 0.20	5.0
0	98.8 ± 0.4	3.72

TABLE XXXI. Resonances in $Si^{28}+p$.

E_p (lab) (Mev)	E_x (P^{29}) (Mev)	Decay [a]	Γ (kev)
(0.703) [b]	(3.40)	γ^a	
(0.944) [b]	(3.64)	γ^a	
(0.989) [b]	(3.68)	γ^a	
1.65 [e]	4.30	p_0 [e] γ_0 [e]	50.5 [f]
2.09 [e]	4.74	p_0 [e] γ_0 [e]	12 [f]
2.27 [f]	4.92	p_1 [f]	<4 [f]
2.64 [d]	5.27	p_1 [d]	
2.9 [e]	5.5	p_0 [e]	broad [e]
3.11 [c]	5.72	p_0 [e] p_1 [c, f]	12 [c]
3.35 [c]	5.96	p_0 [e] p_1 [c]	11 [c]
3.58 [c]	6.18	p_1 [c]	70 [c]
3.71 [c]	6.30	p_1 [c]	40 [c]
3.88 [d]	6.47	p_1 [d]	
3.93 [d]	6.52	p_1 [d]	
3.97 [d]	6.56	p_1 [d]	
4.25 [c]	6.83	p_1 [c]	22 [c]
4.44 [c]	7.01	p_1 [c]	~100 [c] (double?)
	7.25 [c]	p_1 [c]	
	7.44 [c]	p_1 [c]	
4.93 [d]	7.48	p_1 [d]	

[a] Symbols p_0, p_1, γ, and γ_0 refer to elastic proton scattering, inelastic scattering leading to Si^{28} (1), proton capture, and proton capture with a strong ground-state γ transition, respectively.
[b] Se 55. [c] Co 55a. [d] Wi 56. [e] Vo 57. [f] Ne 57.

retical discussion of β^+- and γ-transition probabilities, Go 56b.

II. (a) $Si^{28}(p,\gamma)P^{29}$ $E_b=2.724$
(b) $Si^{28}(p,p'\gamma)Si^{28}$ $E_b=2.724$ $E_x=1.78$

With natural silicon, resonances in yield of γ rays observed for proton energies below 1.2 Mev (Se 55); most of these can be assigned to Si^{29} or Si^{30}; assignment of resonances at 703, 944, and 989 kev uncertain. Resonances at $E_p=1.63$ and 2.08 Mev assigned to $Si^{28}(p,\gamma)P^{29}$. First of these has $\Gamma=50.5$ kev, $\Gamma_\gamma=1.73$ ev, $J=3/2^-$, and the main γ transition goes directly to ground state. Reduced width indicates this is $p_{3/2}$ single particle state (Ne 57).

Yield of 1.78-Mev γ rays from $Si^{28}(p,p'\gamma)Si^{28}$ reaction shows several resonances, Table XXXI. Resonance at $E_p=4.44$ Mev is possibly a doublet (Co 55a).

III. $Si^{28}(d,n)P^{29}$ $Q_m=0.499$

Ground-state Q values: -0.80 ± 0.10 Mev (Pe 48), 0.29 ± 0.04 Mev (Ma 52), and 0.6 ± 0.1 Mev (Ca 57).

Angular distribution measurements at $E_d=9$ Mev yield a $l_p=0$ assignment for ground-state transitions. Levels in P^{29} at 1.30 Mev ($l_p=2$), 1.92 Mev ($l_p=2$), and 3.5 Mev (Ca 57), while additional levels at $E_d=2.5$ and 2.9 Mev in Gr 55f. Also Ca 56a.

IV. $Si^{29}(p,n)P^{29}$ $Q_m=-5.750$

Observed (En 54a, Ty 54).

V. $S^{32}(p,\alpha)P^{29}$ $Q_m=-4.197$

Not observed.

Si^{30}

(Fig. 17)

I. $Mg^{26}(\alpha,p)Al^{29}$ $E_b=10.626$ $Q_m=-3.18$

Resonances, En 54a.

II. $Al^{27}(\alpha,p)Si^{30}$ $Q_m=2.376$

Investigations with α particles from radioactive sources, En 54a.

Nuclear emulsion work at $E_\alpha=8$ Mev yields $Q=2.38\pm0.03$ Mev, and levels in Si^{30} at $E_x=2.23\pm0.02$, 3.52 ± 0.02, 3.80 ± 0.04, 4.83 ± 0.02, 5.28 ± 0.02, 5.52 ± 0.03, 5.94 ± 0.04, 6.52 ± 0.03, 7.10 ± 0.03, and 7.38 ± 0.05 Mev. No levels at 5.07 and 5.62 Mev (Ha 56e). Angular distributions at $E_\alpha=8$ Mev (Vo 56). At $E_\alpha=22$ Mev, levels found from Al absorption at 5.61, 7.35, 8.37, 9.43, 10.04, and 11.03 Mev (Br 49).

Energies of γ rays in coincidence with selected proton groups measured by scintillation spectrometer at $E_\alpha=7.8$ Mev. A 2.32 ± 0.05-Mev γ ray, de-exciting first excited state, and a 3.63 ± 0.15-Mev γ ray, representing unresolved ground-state transitions from second and third excited states. The 5.07-Mev level is de-excited through third level with γ rays of 1.28 ± 0.06 and 3.66 ± 0.15 Mev, and for less than 15% through first level (Al 55). Also La 51, Be 48, Br 55b.

III. $Si^{29}(n,\gamma)Si^{30}$ $Q_m=10.609$

Cross section, Hu 55c.

Energy of γ ray corresponding to ground-state transi-

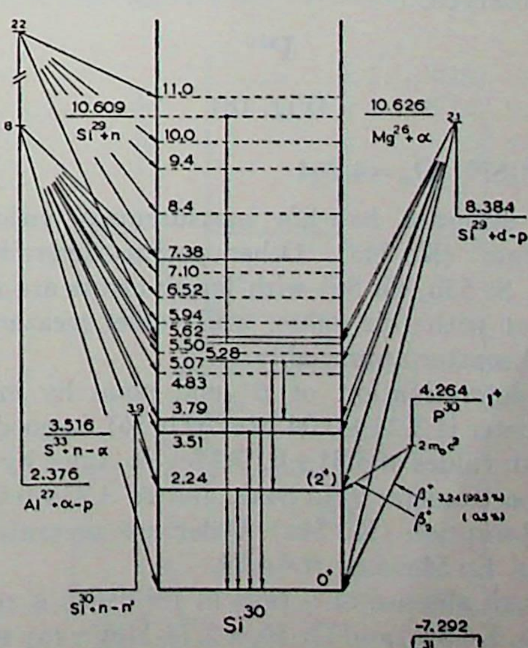


FIG. 17.

TABLE XXXII. The $Si^{29}(d,p)Si^{30}$ reaction.

Q value (Va 52) in Mev	Si^{30} level (Va 52) in Mev	Si^{30} level (Mo 50) in Mev
8.388±0.013	0	0
6.149±0.015	2.239±0.020	(2.4 ±0.2)
4.873±0.009	3.515±0.016	
4.602±0.015	3.786±0.020	3.91±0.15
	(5.075±0.015)	5.00±0.15
	(5.497±0.015)	
	(5.622±0.015)	5.7 ±0.2

tion after capture of thermal neutrons in Si^{29} by pair spectrometer is 10.599±0.011 Mev (Table XXVII). Transition energy obtained by adding 2 kev, nuclear recoil energy, to γ energy (Ki 53c). Same γ ray reported in Ad 56a (Table XXVII). It is uncertain which of the other γ rays given in Ad 56a should be assigned to $Si^{29}(n,\gamma)$ reaction.

IV. $Si^{29}(d,p)Si^{30}$ $Q_m=8.384$

Several proton groups from enriched Si^{29} targets found by high-resolution magnetic analysis ($\theta=90°$) at deuteron energies up to 2.1 Mev (Va 52). Q values and corresponding Si^{30} levels in Table XXXII together with Si^{30} levels obtained by Al absorption at $E_d=3.7$ Mev from enriched targets (Mo 50).

V. $Si^{30}(n,n'\gamma)Si^{30}$

Yield of 2.19-Mev γ ray from natural Si targets measured up to $E_n=3.9$ Mev (Da 56).

VI. $P^{30}(\beta^+)Si^{30}$ See P^{30}

VII. $P^{31}(\gamma,p)Si^{30}$ $Q_m=-7.292$

Not observed.

VIII. $S^{33}(n,\alpha)Si^{30}$ $Q_m=3.516$

Not observed.

P^{30}

(Fig. 18)

I. $P^{30}(\beta^+)Si^{30}$ $Q_m=4.264$

Most accurate half-life measurement yields 2.55 ±0.02 min (Ko 54a). Other recent determinations (Ba 52a, St 53b, Gr 56) with larger errors are in good agreement with this value, while older measurements (En 54a) scatter appreciably.

Best determination of β^+ end point by magnetic spectrometer is 3.24±0.04 Mev (Gr 56), in good agreement with values of 3.31±0.07 Mev obtained by scintillation spectrometer (Hu 54a), and of 3.23±0.07 Mev by Al absorption (Ko 54a). Older less accurate measurements, En 54a. Log $ft=4.84$.

Although absence of γ rays in P^{30} decay is reported in St 53b, Ko 54a, and Gr 56, a 2.24-Mev γ ray recently observed with 0.5% intensity (Mo 56d). Log $ft=4.9$. This uniquely fixes P^{30} ground-state spin as 1^+, if spin of Si^{30} first level is 2^+.

II. $Al^{27}(\alpha,n)P^{30}$ $Q_m=-2.671$

Threshold reported as $E_\alpha=3.409$ Mev corresponding to a Q value of -2.969 Mev (Bu 56c). This leads to a P^{30} mass about 300 kev higher than value from other reactions. Also En 54a.

III. $Si^{28}(He^3,p)P^{30}$ $Q_m=6.318$

Observed, En 54a.

IV. $Si^{29}(p,\gamma)P^{30}$ $Q_m=5.562$

Resonance energies and corresponding P^{30} excitation energies, resonance strengths, and assignments of spins, parities, and isobaric spins in Table XXXIII. The γ-ray energies measured at four lowest resonances lead to Q value of 5.57±0.03 Mev (Va 57). From relative γ-ray intensities reported in Br 56h, Va 57, branching ratios can be computed for resonance levels and several lower P^{30} levels, Fig. 18. Excitation energy of first P^{30} level, which is strongly excited at the 414-kev resonance, is measured by scintillation spectrometer as 688±7 kev (En 54b), and as 690±10 kev (Br 56a). The $T=1$ analog of Si^{30} ground state is expected at about this position (Mo 54, Wi 56a). Isotropy of 690-kev γ ray, together with anisotropy of γ-ray feeding level (1) at 414-kev resonance, presents a strong argument for a

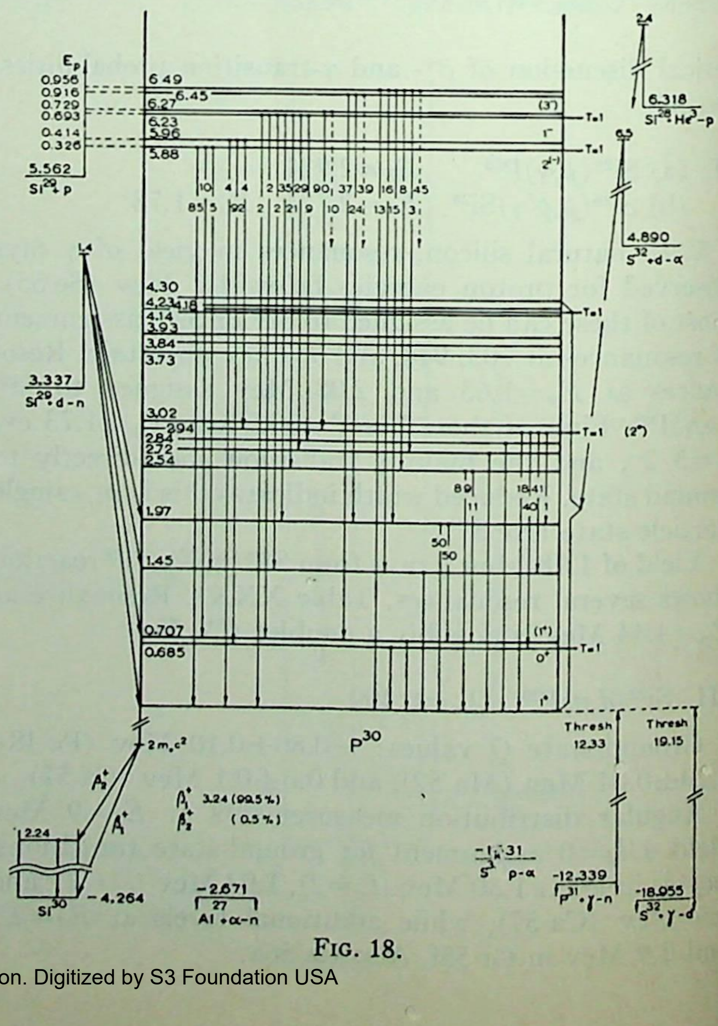


FIG. 18.

TABLE XXXIII. Resonances in $Si^{29}(p,\gamma)P^{30}$.

E_p in kev	E_x (P^{30}) in Mev	$(2J+1)\Gamma_p\Gamma_\gamma/(\Gamma_p+\Gamma_\gamma)$ in ev	J, π, T [g]
326 [a, b, c]	5.877	0.05 [f]	$J=2^-$, $T=1$
414 [a, b, c, d]	5.962	0.23 [e]	$J=1^-$, $T=0$
693 [b, c, d]	6.232	0.11 [e]	(π negative), $T=1$
729 [b, c, d]	6.267	0.11 [e]	$J=(3^-)$, $T=0$
916 [c, d]	6.447		
956 [c, d]	6.486		

[a] Ta 46. [b] Kl 55a. [c] Mi 55. [d] Br 56i. [e] Br 56h. [f] Va 57. [g] All J, π, and T assignments, partly from intensity considerations (Br 56h), and partly from angular distribution measurements (Br 56a, Va 57) are preliminary.

$J=0^+$ assignment to level (1) (Br 56a). Recently observed that level (1) is a close doublet. At 414- and 723-kev resonances, lowest energy γ ray is measured as 686 ± 4 and 686 ± 6 kev, respectively. At 326- and 696-kev resonances, however, the γ-ray energy is distinctly higher, 703 ± 6 and 705 ± 5 kev. Angular distribution of this γ ray, measured at 326-kev resonance, is anisotropic (Va 57). Both levels also observed from $S^{32}(d,\alpha)P^{30}$ reaction (Reaction XI). A preliminary spin assignment to 0.71-Mev $T=0$ level is $J=1^+$ (Va 57).

At least the four lowest resonances seem to decay almost exclusively to either $T=0$ to $T=1$ levels, pointing to negative parity of the resonance levels and operation of E1 isobaric spin selection rule. Remarkable is also strong E2 transition between the lowest two $T=1$ levels.

V. $Si^{29}(p,p'\gamma)Si^{29}$ $E_b=5.562$

In $E_p=2.5$–3.0-Mev region, resonances in yield of a 1.28-Mev γ ray observed at $E_p=2.614$, 2.678, 2.700, 2.720, 2.760, 2.798, 2.843, 2.876, 2.922, 2.934, 2.958, 2.982, and 2.995 Mev. A 2.03-Mev γ ray also observed at 2.678-, 2.700-, 2.798-, 2.876-, 2.934-, and 2.958-Mev resonances. A 2.43-Mev γ ray only appears weakly at 2.982-Mev resonance (Br 57). For γ-ray angular distributions, Si^{29}.

VI. $Si^{29}(d,n)P^{30}$ $Q_m=3.337$

By nuclear emulsion technique, ground-state $Q=3.27\pm0.04$ Mev, and levels found at 0.75 ± 0.06, 1.46 ± 0.06, and 2.00 ± 0.06 Mev by bombarding enriched silicon targets at $E_d=1.4$ Mev (Ma 52). A ground-state Q value of 3.38 ± 0.17 Mev and a level at 1.27 ± 0.48 Mev by bombarding natural silicon targets at $E_d=3.7$ Mev (Pe 58).

VII. $Si^{30}(p,n)P^{30}$ $Q_m=-5.047$

Observed, En 54a.

VIII. $P^{31}(\gamma,n)P^{30}$ $Q_m=-12.339$

Threshold is 12.35 ± 0.2 Mev (Mo 49), 12.4 ± 0.2 Mev (Ka 51a), 2.05 ± 0.20 Mev (Sh 51a), and 12.33 ±0.05 Mev (Sc 55a). Cross section, En 54a, Na 54. Sharp bends in P^{30} yield curve, not corresponding to known levels in P^{30}, observed at 12.58 ± 0.07, 12.75 ±0.08, 12.90 ± 0.08, 13.18 ± 0.10, and 13.38 ± 0.10 Mev (Ba 55).

IX. $S^{32}(\gamma,d)P^{30}$ $Q_m=-18.955$

Threshold reported at 19.15 ± 0.20 Mev (Ka 51a). Yield measurements, En 54a, Ri 55.

X. $S^{32}(n,t)P^{30}$ $Q_m=-12.697$

Observed, En 54a.

XI. $S^{32}(d,\alpha)P^{30}$ $Q_m=4.890$

Magnetic analysis at $E_d=6.0$ and 6.5 Mev and at angles of 90° and 130°, ground-state Q value $=4.887$ ±0.010 Mev, while levels in P^{30} observed at 0.680 ±0.010, 0.708 ± 0.008, 1.451, 1.972, 2.538, 2.723, 2.839, 2.937, 3.018, 3.734, 3.836, 3.926, 4.141, 4.181, 4.230, 4.296, 4.342, 4.421, 4.501, 4.625, 4.734, 4.929, 5.024, 5.200, (5.233), 5.412, 5.504, 5.598, 5.700, and 5.790, all ±0.010 Mev. Levels above 4.30 Mev not indicated in Fig. 18. States at 0.680, 2.937, 4.181, and 4.501 Mev regarded as $T=1$ levels because of low intensity of corresponding groups. They are analogs of Si^{30} ground state and Si^{30} levels at 2.24, 3.51, and 3.79 Mev (Pa 57). A ground-state Q value of 4.831 ± 0.013 Mev, in bad agreement with the value quoted, also reported, in addition to α-particle groups corresponding to levels at 0.693, 1.440, and 1.97 Mev (Le 56b).

XII. $S^{33}(p,\alpha)P^{30}$ $Q_m=-1.531$

Not observed.

Si^{31}

(Fig. 19)

I. $Si^{31}(\beta^-)P^{31}$ $Q_m=1.481$

Half-life, averaged from five determinations (En 54a) in good mutual agreement, 157.2 ± 0.4 min.

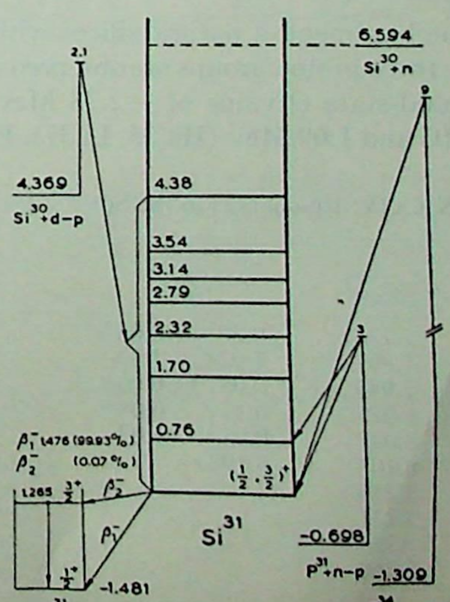


FIG. 19.

The β^- spectrum is simple and has allowed shape. End point measured by lens spectrometer as 1.471 $\pm$0.008 Mev (Mo 52a) and 1.486$\pm$0.012 Mev (Wa 52, Wa 53a). Log ft=5.5. A 1.26-Mev γ ray of intensity 0.07% detected by scintillation spectrometer. Corresponding β^- transition is allowed (log ft=5.6) (Ly 54).

Theoretical remarks concerning Si^{31} decay, Go 56b.

II. $Si^{30}(n,\gamma)Si^{31}$ $Q_m=6.594$

Cross section, Hu 55c.

III. $Si^{30}(d,p)Si^{31}$ $Q_m=4.369$

By magnetic analysis at deuteron energies up to 2.1 Mev, ground-state Q value=4.364$\pm$0.007 Mev, and levels in Si^{31} found at 0.757$\pm$0.007, 1.699$\pm$0.007, 2.319$\pm$0.008, 2.791$\pm$0.008, 3.140$\pm$0.008, 3.539$\pm$0.008, and 4.384$\pm$0.008 Mev (Va 52). Also En 54a.

IV. $P^{31}(n,p)Si^{31}$ $Q_m=-0.698$

By measuring pulse height from a P_2O_3 gas-filled ionization chamber bombarded by $D(d,n)$ neutrons, ground-state Q value=-0.97 ± 0.13 Mev and a level in Si^{31} at 0.7 Mev (Me 48). Cross section, En 54a. Resonances under P^{32}.

V. $S^{34}(n,\alpha)Si^{31}$ $Q_m=-1.309$

Cross section, En 54a.

P^{31}

(Figs. 20 and 21)

I. (a) $Al^{27}(\alpha,n)P^{30}$ $E_b=9.668$ $Q_m=-2.722$
(b) $Al^{27}(\alpha,p)Si^{30}$ $E_b=9.668$ $Q_m=2.376$

Thick target resonances observed with radioactive α-particle sources, En 54a.

II. $Si^{28}(\alpha,p)P^{31}$ $Q_m=-1.920$

From bombardment of natural silicon with Th$(B+C)$ α particles, three proton groups are observed corresponding to ground-state Q value of -2.23 Mev and levels in P^{31} at 1.05 and 1.69 Mev (Ha 35, Li 37). From 7-Mev cyclotron α particles on natural silicon, a γ ray is found of 2.3$\pm$0.3 Mev as measured by absorption (Al 48).

TABLE XXXIV. Resonances in the $Si^{30}(p,\gamma)P^{31}$ reaction.

E_p (kev)	E_x (P^{30}) (Mev)	$(2J+1)\Gamma\gamma$ (ev)		J, π
500 [a]	7.776	0.20 [a]	0.06 [b]	3/2+ [a, b]
625 [a]	7.897	1.92 [a]	1.8 [b]	1/2 [a, b]
675 [a]	7.945	0.087 [a]	0.06 [b]	3/2+ [a, b]
760 [a]	8.028	0.17 [a]	0.08 [b]	
775 [a]	8.042	0.93 [a]	1.0 [b]	3/2+ [a, b]
840 [a]	8.105	0.192 [a]		(3/2, 5/2)+ [a]
955 [a]	8.216			
980 [a]	8.241			
995 [a]	8.255			
1000 [a]	8.260			

[a] Br 56i.
[b] Ho 57.

III. $Si^{30}(p,\gamma)P^{31}$ $Q_m=7.292$

Resonances observed, corresponding P^{31} excitation energies, resonance strengths, and resonance spins and parities deduced from γ-ray angular distribution measurements (Br 56i) in Table XXXIV. Proton energies at five lower resonances in fair agreement with Kl 55a. Also Se 55. A resonance reported at E_p=367 kev (Ta 46) not observed in later work (Kl 55a). Twenty-five resonances reported in the 1.7–2.3-Mev region. Those at E_p=1.73 and 1.82 Mev have $J=3/2^+$ (Pa 56).

Gamma-ray intensities at five lower resonances in En 56b, Ho 57. Corresponding branching percentages in Fig. 21. Branching percentages indicated for 3.414-, 4.188-, 4.430-, 5.012-, 6.22-, 6.43-, and 6.55-Mev levels and for resonances at E_p=840, 980, and 1000 kev from Br 56i.

Gamma-ray angular distributions (Br 56i, En 56b), and γ-γ angular correlation measurements (Pa 56, Li 57) in agreement with a $J=1/2^+$ assignment to P^{31}

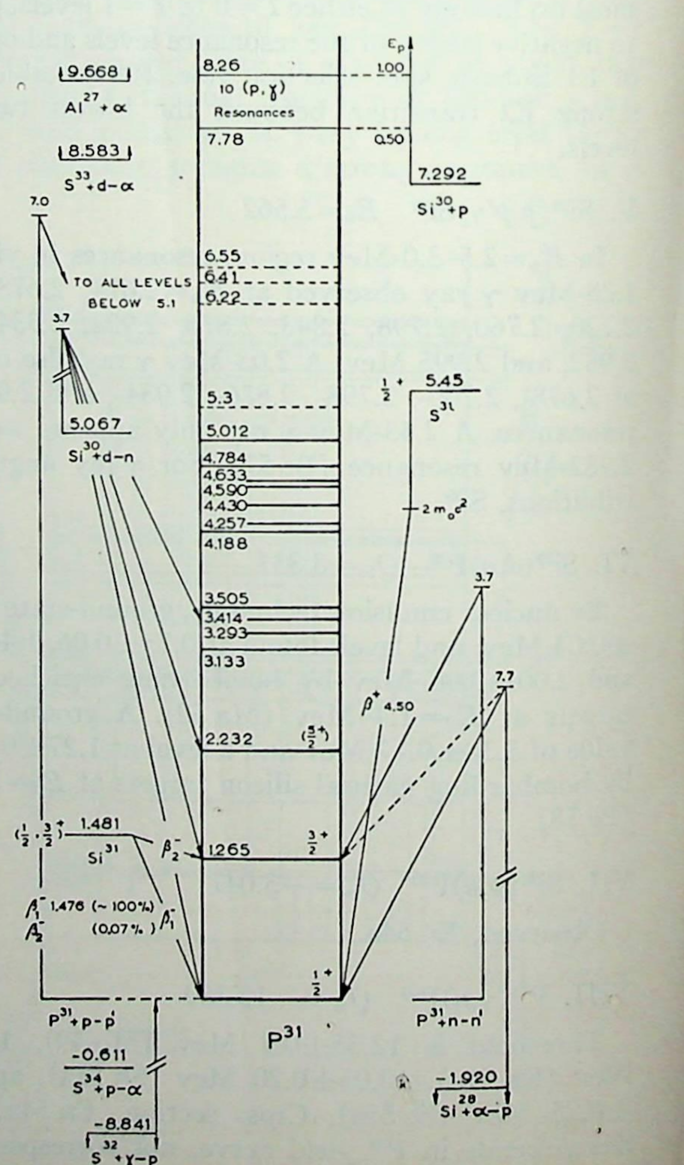
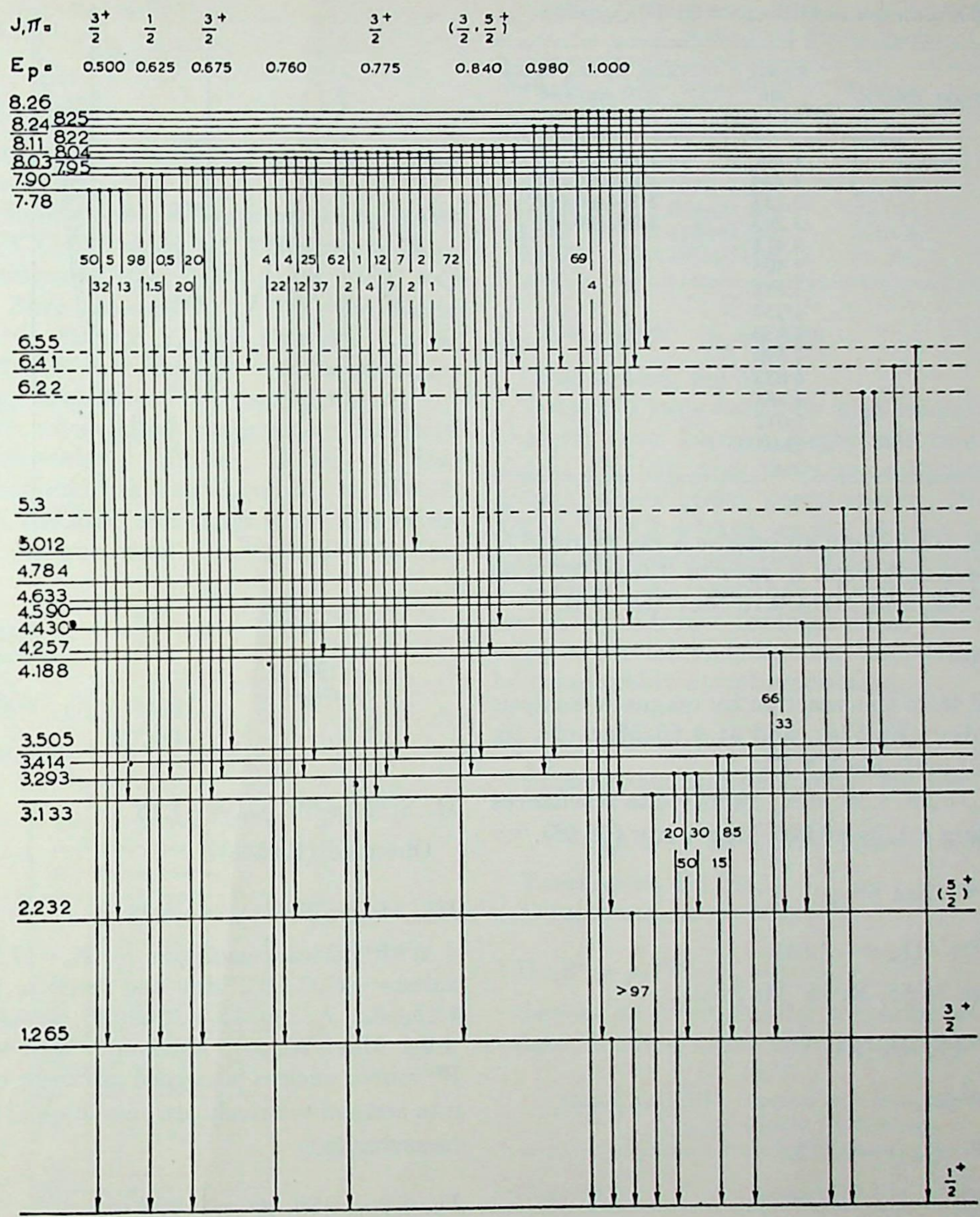


FIG. 20.



Gamma transitions in P^{31}

FIG. 21.

ground state, and indicate $J=3/2^+$ for level (1) and very probably $J=5/2^+$ for level (2). The latter level decays for at least 97% by an E2 transition to ground state (Li 57), in remarkable analogy with Si^{29}. Collective motion might be responsible for these anomalies.

Positive parities assigned to resonance levels in Table XXXIV found from γ-ray angular distribution measurements which can only be explained through mixing of M1 and E2 radiation.

IV. $Si^{30}(d,n)P^{31}$ $Q_m=5.067$

From natural silicon bombardments at $E_d=3.7$ Mev and with nuclear emulsion technique for neutron detection, Q values are found of 4.56 ± 0.14, 4.12 ± 0.16, 3.57 ± 0.15, and 2.78 ± 0.16 Mev (Pe 48). With enriched Si^{30} targets and nuclear emulsions, at $E_d=1.4$ Mev Q values have been measured of 4.92, 4.59, 3.73, 2.70, and 1.51 Mev, all ±0.04 Mev (Ma 52). Averaging these results (and assuming that in the earlier data ground-state transition has been missed) following excitation energies in P^{31} are obtained: 0.4, (0.9), 1.4, 2.3, and 3.4 Mev.

V. $Si^{31}(\beta^-)P^{31}$ See Si^{31}

VI. $P^{31}(n,n'\gamma)P^{31}$

No γ rays observed at $E_n=1.2$ Mev (Va 56a). At $E_n=2.45$ Mev, a γ ray of 1.25 ± 0.03 Mev is reported

TABLE XXXV. Levels in P^{31} found from $P^{31}(p,p')P^{31}$.

Reference:	En 57a	Va 56, Va 57c, Va 57a
Incident proton energy (Mev):	7.04	3.72 and 4.66
Angle:	90° and 130°	90°
Level in P^{31} (Mev):	1.267	1.264±0.004
	2.234	2.230±0.005
	3.133	3.134±0.006
	3.293	3.292±0.005
	3.414	
	3.505	
	4.188	
	4.257	
	4.430	
	4.590	
	4.633	
	4.784	
	5.012	
	all ±0.005	

(Cr 56a). At $E_n=3.5$ Mev, an intense γ ray is found of 1.24±0.05 Mev, and weak γ rays of 1.0±0.05, 1.60 ±0.07, 1.75±0.08, and 2.05±0.18 Mev (Sc 54d).

VII. $P^{31}(p,p')P^{31}$

Levels found from this reaction by magnetic analysis at $E_p=7.04$ Mev (En 57a), and at 4.65 Mev (Va 56, Va 57a, Va 57c) in Table XXXV.

In region $E_p=2.3$–3.36 Mev, twenty-one resonances observed yielding a 1.26±0.015-Mev γ ray (Ol 55).

VIII. $S^{31}(\beta^+)P^{31}$ See S^{31}

IX. $S^{32}(\gamma,p)P^{31}$ $Q_m=-8.841$

Cross section, Jo 55, Ri 55, Mo 55a.

X. $S^{33}(d,\alpha)P^{31}$ $Q_m=8.583$

Not observed.

XI. $S^{34}(p,\alpha)P^{31}$ $Q_m=-0.611$

Not observed.

GENERAL REMARKS

Theoretical remarks concerning P^{31} ground-state configuration, Hu 55b, Go 56b.

S^{31}

(Fig. 22)

I. $S^{31}(\beta^+)P^{31}$ $Q_m=5.45$

Half-life averaged from two accurate post-war determinations (Ha 52a, Hu 54) is 2.62±0.09 sec. The β^+ end point determined with cloud chambers as 3.85 ±0.07 Mev (Wh 41) and 3.87±0.15 Mev (El 41), and with a scintillation spectrometer as 4.50±0.10 Mev (Hu 54). Log $ft=3.7$, indicating a superallowed transition. Hence, S^{31} has $J=1/2^+$.

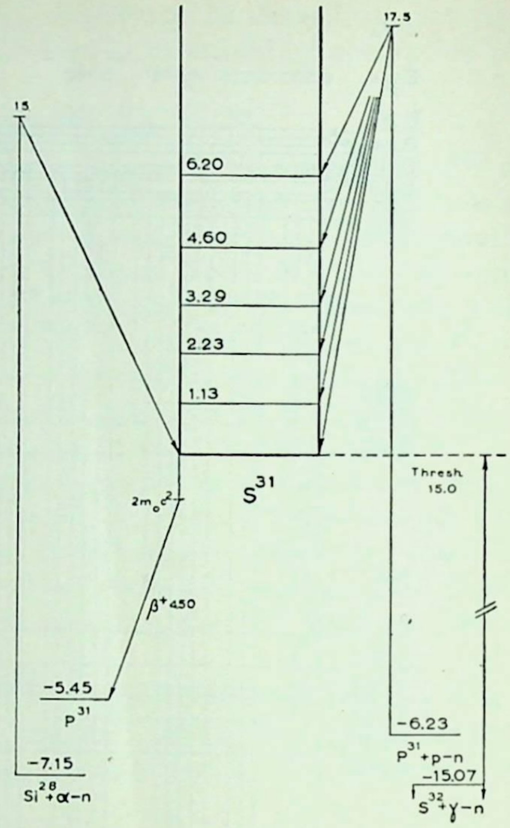


FIG. 22.

II. $Si^{28}(\alpha,n)S^{31}$ $Q_m=-7.15$

Observed (En 54a).

III. $P^{31}(p,n)S^{31}$ $Q_m=-6.23$

With nuclear emulsions at $E_p=17.5$ Mev, the Q value $=-6.03\pm0.2$ Mev and levels in S^{31} observed at 1.13±0.2, 2.23±0.15, 3.29±0.15, 4.60±0.15, and 6.20 ±0.2 Mev (Ru 55). Agreement with level scheme of P^{31} mirror nucleus is as good as can be expected taking into account relatively low resolution of neutron energy measurement.

IV. $S^{32}(\gamma,n)S^{31}$ $Q_m=-15.07$

Threshold measured as 15.0±0.3 Mev (Be 47a), 14.8±0.4 Mev (Mc 49), and 15.0±0.1 Mev (Ha 52a). Cross section, En 54a, Mo 55a.

Si^{32}

(not illustrated)

I. $Si^{32}(\beta^-)P^{32}$ $Q_m=0.10$

Radioactive Si^{32} produced from reaction $Cl^{37}(p,\alpha 2p)Si^{32}$ at $E_p=340$ Mev. From measured activity and estimated reaction cross section, a half-life is calculated between 100 and 710 years. The β^- spectrum end point is ~100 kev. There are no γ rays (Li 53a). Also Li 53. Also produced from neutron capture in Si^{31} in a high-flux reactor. From yield, half-life (in years) computed as 600 times the cross section (in barns) (Tu 54).

P^{32}

(Fig. 23)

I. $P^{32}(\beta^-)S^{32}$ $Q_m = 1.704$

Six determinations (En 54a) of half-life yield average value of 14.32 ± 0.03 days.

Determinations of β^- end point by magnetic spectrometer in Table XXXVI. Average is 1.7080 ± 0.0026 Mev. Spectrum has allowed shape. Coefficient $r < 0.03$ mc^2 in a possible Fierz correction $(1+r/W)^{\frac{1}{2}}$ to Kurie plot (Po 56). No discrete γ rays observed (La 54, Go 54b). Log $ft = 7.9$, very large for an allowed transition but this can be explained by l forbiddenness.

Many authors investigated continuous γ radiation (internal bremsstrahlung) (En 54a, Go 54b, Mi 54a, Li 55, Fo 56), internal pair formation (Mc 54, We 54, Mi 56, Gr 56b, Hu 56a), and other small effects involving atomic electron cloud (Ch 54, Ch 55a, Re 55).

II. $Si^{29}(\alpha,p)P^{32}$ $Q_m = -2.474$

Observed, En 54a.

III. $Si^{30}(He^3,p)P^{32}$ $Q_m = 7.494$

Observed, En 54a.

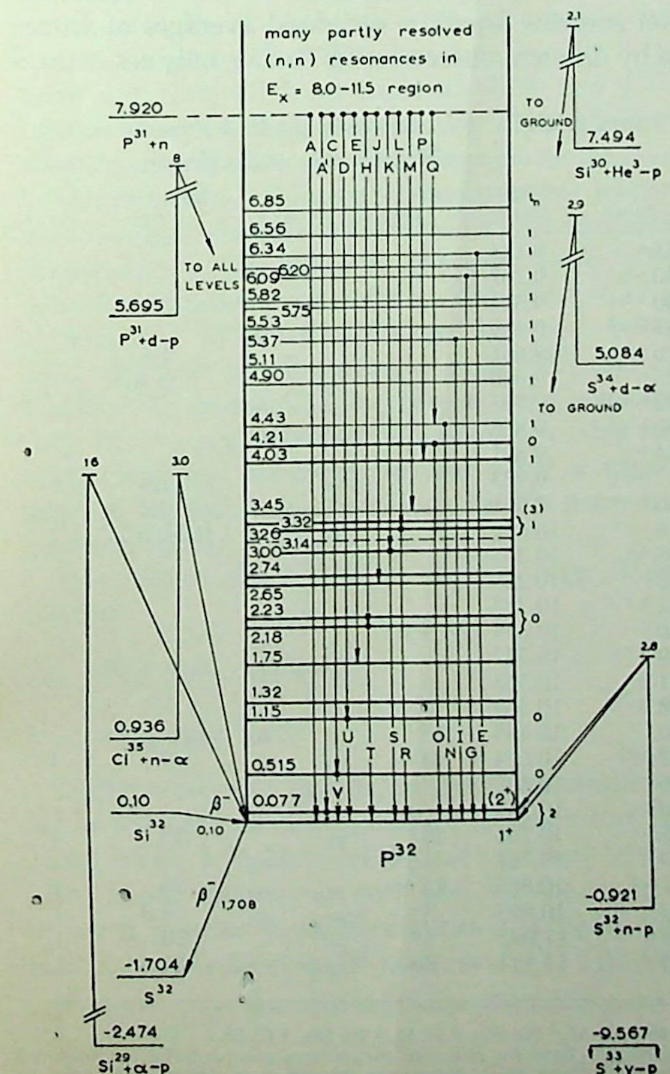


FIG. 23.

TABLE XXXVI. Magnetic-spectrometer determinations of the $P^{32}(\beta^-)S^{32}$ end point.

Reference	End point (Mev)	Reference	End point (Mev)
Ly 37	1.69 ± 0.03	Sh 51	1.695 ± 0.005
La 39a	1.72 ± 0.01	Je 52	1.704 ± 0.008
Wi 41	1.75 ± 0.02	Mo 52a	1.697 ± 0.010
Si 46	1.712 ± 0.008	Wo 54	1.714 ± 0.008
La 49a	1.689 ± 0.010	An 54a	1.712 ± 0.008
Ag 50	1.718 ± 0.010	Da 54a	1.711 ± 0.006
Wa 50	1.708 ± 0.008	Po 56	1.712 ± 0.004

IV. $P^{31}(n,\gamma)P^{32}$ $Q_m = 7.920$

Cross section, Hu 55c.

Twenty γ rays found by high-resolution pair spectrometry from capture of thermal neutrons in phosphorus (Ki 52), and three more found with a two-crystal scintillation spectrometer (Br 56e). Table XXXVII. All γ rays, except B, can be regarded as transitions from capturing state to known levels in P^{32}, or as transitions from known levels to ground. However, intensities of γ rays feeding and de-exciting a level may differ appreciably, indicating that decay scheme may be considerably more complicated.

V. $P^{31}(n,n)P^{31}$ $E_b = 7.920$

Cross section and resonances, Hu 55c, Pa 55b, Pa 55.

VI. $P^{31}(n,p)Si^{31}$ $E_b = 7.920$ $Q_m = -0.698$

Resonances, En 54a.
Q-value measurements, Si^{31}.

VII. $P^{31}(d,p)P^{32}$ $Q_m = 5.695$

Levels in P^{32} found by magnetic analysis, and l_n values deduced from angular distributions of corre-

TABLE XXXVII. Gamma rays from thermal neutron capture in phosphorus.

γ ray [a]	Energy in Mev	Intensity in photons per 100 captures
A	7.94 ± 0.03	0.5
A'	7.85 ± 0.05	1.5
B	7.62 ± 0.03	2
C	7.42 ± 0.03	7
D	6.76 ± 0.03	24
E	6.33 ± 0.03	0.7
F	6.14 ± 0.03	1.5
G	6.02 ± 0.04	1.5
H	5.71 ± 0.03	6
I	5.41 ± 0.03	2
J	5.27 ± 0.03	8
K	4.92 ± 0.03	4
L	4.68 ± 0.03	26
M	4.49 ± 0.03	4
N	4.38 ± 0.03	11
O	4.20 ± 0.03	7
P	3.92 ± 0.03	25
Q	3.55 ± 0.03	22
R	3.28 ± 0.04	9
S	3.04 ± 0.04	7
T	2.19 ± 0.03	41
U	1.13 ± 0.03	14
V	0.51 ± 0.02	28

[a] [illegible] S, Br [illegible] for γ rays T through V.

TABLE XXXVIII. Levels in P^{32} from the reaction $P^{31}(d,p)P^{32}$.

Reference: E_d (Mev):	Va 52b 1.8 and 2.0	Da 57 8	
	E_x (P^{32}) (Mev)	E_x (P^{32}) (Mev)	l_n
	0	0	} 2
	0.077±0.002		
	0.515±0.005		0
	1.154±0.007		0
	1.316±0.008		...[a]
	1.750±0.009		...[a]
	2.177±0.009		} 0 or 1
	2.227±0.009		
	2.650±0.008		...[a]
	2.742±0.008		...[a]
	2.999+0.010		...[a]
	(3.141±0.012)		...[a]
	3.259±0.009		} 1
	3.318±0.009		
		3.45±0.1	(3)
	4.032±0.009		1
	4.207±0.010		0
		4.43±0.1	1
		4.66±0.1	weak
		4.90±0.07	1
		5.11±0.1	(isotropic)
		5.37±0.07	1
		5.53±0.07	1
		5.75±0.1	weak
		5.82±0.07	1
		6.09±0.07	1
		6.20±0.1	weak
		6.34±0.1	1
		6.56±0.07	1 or 2
		6.69±0.1	weak
		6.85±0.1	weak

[a] Not observed or weak.

sponding proton groups in Table XXXVIII. Ground-state Q value is 5.704±0.008 Mev (Va 52b). The $l_n=2$ assignment to ground-state doublet also found at $E_d=7.2$ Mev (Pa 52) and at $E_d=14.3$ Mev (Bl 53). The $l_n=0$ admixture to this group is less than 5% (Pa 52).

Experimental yield ratio (excited-state protons over ground-state protons) at $\theta=90°$ is 1.7 at $E_d=1.8$ Mev and 1.2 at $E_d=2.0$ Mev (Va 52b). This agrees reasonably with predicted ratio of 1.67 for ground-state spin $J=1^+$ and excited-state spin $J=2^+$ (En 53a, En 54).

See S^{32}, reaction V, for γ rays from $P^{31}+d$ reactions.

VIII. $S^{32}(n,p)P^{32}$ $Q_m=-0.921$

Ionization chamber measurement with pulse-height analysis yields -0.93 ± 0.1 Mev for ground-state Q value (Hu 41). At $E_n=2.56$ Mev, a 77±2-kev γ ray observed with a scintillation spectrometer (Da 56c).

Cross section, En 54a, Hu 55a.

Resonances, S^{33}.

IX. $S^{33}(\gamma,p)P^{32}$ $Q_m=-9.567$

Not observed.

X. $S^{34}(\vec{d},\alpha)P^{32}$ $Q_m=5.084$

By magnetic analysis at $E_d=1.6$–2.9 Mev and $\theta=63°$ and 90°, with enriched S^{34} targets, ground-state $Q=5.04$ ±0.02 Mev (Le 56b). This might be too small since also the value obtained by these authors for $S^{32}(d,\alpha)P^{30}$ ground-state Q value is too small by 60 kev.

XI. $Cl^{35}(n,\alpha)P^{32}$ $Q_m=0.936$

From ionization chamber pulse-height analysis, ground-state Q values of 1.07±0.15 Mev (Fo 52) and 0.97±0.16 Mev (Ad 53).

Cross section, En 54a.

GENERAL REMARKS

Theoretical discussion of P^{32} spin, Hi 54. Discussion regarding doublet levels in P^{32}, In 53.

S^{32}

(Fig. 24)

I. $Si^{29}(\alpha,n)S^{32}$ $Q_m=-1.553$

Not observed.

II. (a) $P^{31}(p,\gamma)S^{32}$ $Q_m=8.841$
(b) $P^{31}(p,p'\gamma)P^{31}$ $E_b=8.841$
(c) $P^{31}(p,\alpha)Si^{28}$ $E_b=8.841$, $Q_m=1.923$

Resonances in these reactions in Table XXXIX. Proton energies listed are weighted averages of values given by different observers. In Pa 55a, only resonances

TABLE XXXIX. Resonances in the $P^{31}+p$ reactions.

E_p (kev)	E_x (S^{32}) (Mev)	Γ (kev)	Decay [g]	$(2J+1)\Gamma_\gamma$ [e] (ev)	J, π [e]
355 [a]	9.185		γ		
440 [a, b, c]	9.267		γ		
540 [a, b, c]	9.364		γ		
648 [b, c, d]	9.469		γ α_0		
816 [b, c, e]	9.631		γ_1	1.8	
825 [e]	9.640		γ_0	0.36	1
892 [b, c, d]	9.705		γ α_0		
1050 [b, c, d, f]	9.858		γ α_0		
1084 [b, c]	9.891		γ		
1117 [b, c, e]	9.923		γ_0	1.9	1
1146 [b, c, e]	9.951		γ_1	2.2	
1248 [b, c, e]	10.050		γ_0 γ_1	0.24, 3.7 [h]	2
1408 [b, c]	10.205		γ		
1443 [b, c]	10.239		γ		
1482 [b, c]	10.283		γ		
1520 [f]	10.314	9.0 [f]	α_0		
1549 [b, c, e]	10.342		γ_1		
1571 [b, c]	10.363		γ		
1598 [b, c]	10.389		γ		
1640 [f]	10.430	~5 [f]	α_0		
1892 [c, e, f]	10.674	24 [e, f]	γ_0 α_0	37	1^-
1916 [c]	10.697		γ		
1980 [e]	10.759		α_0		
1985 [c, e]	10.764	10 [e]	γ_0	13	1
1990 [e]	10.767		α_0		
2027 [e, f]	10.805	24 [e, f]	γ_0 α_0	26	1^-
2120 [e]	10.895	5 [e]	γ_0 α_0	4.0	1^-
2320 [e]	11.092	8 [e]	γ_0	20	1
2340 [e]	11.111	8 [e]	γ_0	64	1

[a] Ta 46. [b] Gr 51. [c] Ke 56a. [d] Fr 51. [e] Pa 55a. [f] Cl 55.
[g] The symbols have the following meaning: γ gamma radiation observed without specification as to energy, γ_0 transition to S^{32} ground state, γ_1 transition to S^{32} first excited state, α_0 ground-state transition of $P^{31}(p,\alpha)Si^{28}$ reaction. [h] Two values listed relate to γ_0 and γ_1, respectively.

are listed which decay either to S^{32} ground state or to first excited state. List of resonances decaying through the $P^{31}(p,\alpha)Si^{28}$ ground-state transition is not complete. Altogether, there are 16 resonances in $E_p=1.0$–2.5-Mev region, decaying in this way, most of which do not decay by γ-ray emission (Cl 55).

There are 21 resonances in $E_p=2.3$–3.36-Mev region showing a 1.26 ± 0.015-Mev γ ray from $P^{31}(p,p'\gamma)P^{31}$ reaction. For $E_p>2.47$ Mev, also a 1.8-Mev γ ray observed from $P^{31}(p,\alpha)Si^{28}$ leading to first level in Si^{28} (Ol 55).

Spin and parity assignments in Table XXXIX from γ-ray angular distribution measurements. Spin of first level in S^{32} probably 2^+ (Pa 55a). For γ-ray energies measured at several resonances, Ke 56a. Also Sm 54, Va 56. See P^{31} and Si^{28} for levels observed in these nuclei from reactions b and c.

III. $P^{31}(d,n)S^{32}$ $Q_m=6.616$

With nuclear emulsions, ground-state $Q=6.63\pm0.08$ Mev measured at $E_d=8$ Mev. Transitions observed to known lower levels in S^{32} and to new levels at 6.29, 7.28, and 8.33 Mev, all ±0.10 Mev. Angular distribution measurements yield $l_p=0$ for transitions to ground and to 3.78-Mev level, and $l_p=2$ for transitions to 2.24-Mev level (El 55a). This in agreement with results in Ca 55, obtained at $E_d=9$ Mev with a triple ionization chamber for neutron detection. However, transition to 3.78-Mev level not resolved from those to 4.29- and 4.47-Mev levels. Group leading to 5.76-Mev level has an $l_p=2$ angular distribution. Also En 54a.

Gamma rays, following bombardment of phosphorus with 4.6-Mev deuterons, observed with a magnetic pair spectrometer with following energies: 4.43 ± 0.03, 4.73 ± 0.03, 4.95 ± 0.03, 5.31 ± 0.03, 5.81 ± 0.03, 6.13 ± 0.03, 6.86 ± 0.03, 7.49 ± 0.07, 8.19 ± 0.03, and 8.56 ± 0.03 Mev. Neutron binding energies in S^{32}, P^{32}, and Si^{29} are 15.08, 7.92, and 8.47 Mev, respectively, showing that only last two γ rays can be assigned to this reaction with certainty. However, several other γ rays could also be regarded as transitions in S^{32} from level to ground (Be 55).

Theoretical discussion of lowest three S^{32} states, Hu 55b.

IV. $P^{32}(\beta^-)S^{32}$ See P^{32}

V. $S^{32}(e,e')S^{32}$

Differential cross section measured for elastic scattering of 187-Mev electrons, and for inelastic scattering to 2.24-, 3.78-, 5.76-Mev levels and to a new level at 6.6 Mev. From comparison with $\sigma(\theta)$ for the excitation of 2^+ levels in other even-even nuclei it is highly probable that 2.24-Mev level in S^{32} has also $J=2^+$ (He 56).

VI. $S^{32}(n,n')S^{32}$

Inelastic scattering of neutrons studied with scintillation spectrometers. At $E_n=2.56$ Mev, a γ ray of

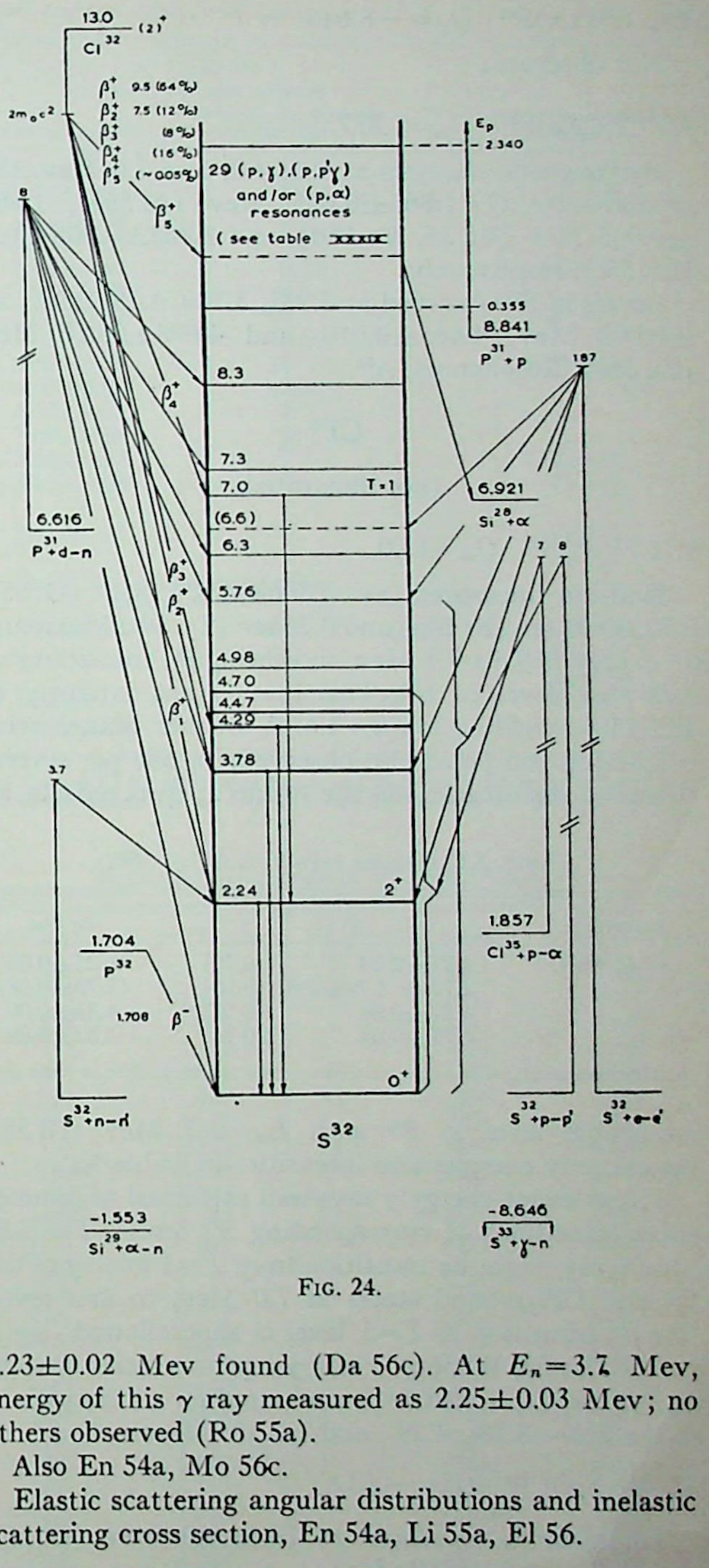


FIG. 24.

2.23±0.02 Mev found (Da 56c). At $E_n=3.7$ Mev, energy of this γ ray measured as 2.25 ± 0.03 Mev; no others observed (Ro 55a).

Also En 54a, Mo 56c.

Elastic scattering angular distributions and inelastic scattering cross section, En 54a, Li 55a, El 56.

VII. $S^{32}(p,p')S^{32}$

By magnetic analysis, at $E_p=8$ Mev, levels in S^{32} found at 2.25, 3.81, 4.32, 4.50, 4.74, 5.04, and 5.83 Mev, all ±0.02 Mev (Ar 52). By comparison with the results obtained by these authors in Al^{27}, excitation energies are probably too high by an amount slowly increasing up to 70 kev for highest levels.

See Cl^{33} for resonances in yield of a 2.2-Mev γ ray.

VIII. $Cl^{32}(\beta^+)S^{32}$ See Cl^{32}

IX. $S^{33}(\gamma,n)S^{32}$ $Q_m=-8.646$

Not observed.

X. $Cl^{35}(p,\alpha)S^{32}$ $Q_m=1.857$

By magnetic analysis at 2.5, 4.0, and 7.0 Mev, the ground-state $Q=1.865\pm0.015$ Mev (Al 55a), 1.860 ±0.005 Mev (Va 56, Va 57d), and 1.863 ± 0.008 Mev (En 56c), respectively.

Levels in S^{32} observed at 2.237, 3.780, 4.287 Mev, all ±0.008 Mev, 4.465 ± 0.010, and 4.698 ± 0.010 Mev (En 56c). Resonances, A^{36}.

Cl^{32}

(not illustrated)

I. $Cl^{32}(\beta^+)S^{32}$ $Q_m=13.0$

Half-life measured as 0.306 ± 0.004 sec (Gl 55), 0.32 ± 0.01 sec (Br 54a) and 0.28 sec (Ty 54). Maximum β^+ energy is 9.5 ± 0.4 Mev, indicating β^+ transitions to 2.24-Mev level in S^{32}. This branch has intensity of $48\pm15\%$, yielding $\log ft=4.6$. A weaker branch with $\sim$7.5-Mev end point also observed. In one per several thousand disintegrations the β^+ decay proceeds to an α-unstable level in S^{32} with $E_\alpha=2$–3 Mev (Gl 55). Gamma-ray energies and intensities in Table XL.

TABLE XL. Gamma rays from $Cl^{32}(\beta^+)S^{32}$.

Reference:	Gl 55		Br 54a
E_γ in Mev	2.21 ± 0.03	(0.70)	2.25 ± 0.04
	2.77 or 3.79 ± 0.08	(0.10)	(3.79 ± 0.08)
	4.27 ± 0.08	(0.07)	4.33 ± 0.09
	4.77 ± 0.04	(0.14)	4.82 ± 0.08

Three lowest-energy γ rays well explained as ground-state transitions of corresponding S^{32} levels. The 4.8-Mev γ ray might be transition from $T=1$ analog of the P^{32} and Cl^{32} ground states at 7.0 Mev, to first level. The β^+ transition to $T=1$ level is superallowed ($\log ft=3.8$) (Bo 55, Wi 56a). From γ-ray intensities branchings of 64, 12, 8, and 16% computed for β^+ transitions to the 2.24-, 3.78-, 4.29-, and 7.0-Mev levels.

II. $S^{32}(p,n)Cl^{32}$ $Q_m=-13.8$

Threshold is measured as 14.3 ± 0.5 Mev (Gl 55), and 14.5 ± 0.6 Mev (Br 54a).

At $E_p=17.5$ Mev, neutron groups leading to Cl^{32} levels at 1.0, 1.4, and 2.0 Mev observed with nuclear emulsions (Aj 55).

P^{33}

(not illustrated)

I. $P^{33}(\beta^-)S^{33}$ $Q_m=0.249$

Average value of 24.5 ± 0.2 days for half-life from four determinations (En 54a, Ni 54).

The β^- end point determined by magnetic spectrometer as 0.27 ± 0.02 Mev (Sh 51c), 0.26 ± 0.02 Mev (Je 52), 0.251 ± 0.005 Mev (El 54), and 0.249 ± 0.002 Mev (Ni 54), and as 0.246 ± 0.005 Mev by Al absorption (We 52). The β^- transition is allowed ($\log ft=5.1$) but l forbidden. No γ rays found (We 52, also La 54).

II. $Si^{30}(\alpha,p)P^{33}$ $Q_m=-2.982$

Not observed.

III. $S^{33}(n,p)P^{33}$ $Q_m=0.534$

Cross section, En 54a, Ni 54.

IV. $S^{34}(\gamma,p)P^{33}$ $Q_m=-10.885$

Observed, En 54a.

V. $S^{36}(p,\alpha)P^{33}$ $Q_m=0.5$

Not observed.

VI. $Cl^{37}(\gamma,\alpha)P^{33}$ $Q_m=-7.864$

Observed, En 54a.

S^{33}

(Fig. 25)

I. $Si^{30}(\alpha,n)S^{33}$ $Q_m=-3.516$

Not observed.

II. $S^{32}(n,\gamma)S^{33}$ $Q_m=8.646$

Cross sections and abundances of sulfur isotopes (Hu 55c) indicate that probably most of thermal-neutron captures in natural sulfur occur in S^{32}.

Gamma rays from thermal neutron capture in natural sulfur in Table XLI. Remarkable are strong E1 transitions to and from p states in S^{33} at 3.22 and 5.71 Mev. Of all proton groups observed from $S^{32}(d,p)S^{33}$ reaction at $E_d=6.55$ Mev, both at 90° and at 130°, groups leading to these levels have also largest intensity (see Reaction V). Several of weaker γ rays of Table XLI might also be explained as transitions to or from known levels in S^{33}, but their assignment is less certain.

III. $S^{32}(n,n)S^{32}$ $E_b=8.646$

Cross section and resonances, Hu 55c, Mo 55b, Pa 55, Mo 56b, La 56a. Fifteen resolved resonances observed below $E_n=1.1$ Mev. Characteristics of seven lowest-energy resonance in Table XLII. Recently, $J\geq5/2$, $l_n\geq2$ reported for 585-kev resonance, La 56a.

IV. (a) $S^{32}(n,p)P^{32}$ $E_b=8.646$ $Q_m=-0.921$
(b) $S^{32}(n,\alpha)Si^{29}$ $E_b=8.646$ $Q_m=1.553$

Cross section, En 54a, Hu 55a.

Also P^{32} and Si^{29} for reactions (a) and (b), respectively.

TABLE XLI. Gamma rays from thermal neutron capture in sulfur.

Reference:	Gr 55a		Ki 52, Ki 54a		
Method:	Magnetic Compton spectrometer		Magnetic pair spectrometer		
Gamma ray	Energy in Mev	Intensity in photons per 100 captures	Energy in Mev	Intensity in photons per 100 captures [a]	Probable transition in S^{33} [b]
A	8.63 ±0.04	1.8	8.64±0.02	1.2	(*c*)→(0)
B	7.80 ±0.04	2.6	7.78±0.03	1.6	(*c*)→(1)
C	7.42 ±0.05	0.6	7.42±0.03	0.3	
D	7.20 ±0.05	0.5	7.19±0.03	0.2	
E	6.62 ±0.05	0.5	6.64±0.03	0.25	(*c*)→(2)
F	5.88 ±0.05	0.4	5.97±0.06	0.6	
G	5.44 ±0.02	48.0	5.43±0.02	60	(*c*)→(7)
H	5.07 ±0.03	3.5	5.03±0.06	3.5	
I	4.87 ±0.02	12.0	4.84±0.06	11	(32)→(1)
J	4.58 ±0.03	1.2	4.60±0.03	3.5	
K	4.40 ±0.02	6.0	4.38±0.03	7	
L	3.66 ±0.05	2.0	3.69±0.05	4	
M	3.41 ±0.01	7.1	3.36±0.05	7	(32)→(3)
N	3.27 ±0.01	19.0	3.21±0.03	20	(7)→(0)
N'	3.10 ±0.03	2.0			
O	2.975±0.01	13.0	2.94±0.05	20	(*c*)→(32)
P	2.82 ±0.02	4.0			
Q	2.70 ±0.05	2	Br 56e. Two-crystal scintillation spectrometer		
R	2.550±0.02	4.4			
S	2.415±0.01	30	2.34±0.03	37	(7)→(1)
T	2.29 ±0.03	5			(3)→(0)
U	2.00 ±0.02	3			(2)→(0)
V			1.52±0.05	1	(3)→(1)
W	0.84 ±0.01	47	0.84±0.02	56	(1)→(0)

[a] Where possible, intensities are taken from Ki 54a. For γ rays *C, D, F, H*, and *J*, from Ki 52, multiplied by correction factor varying from 0.4 to 0.7.
[b] Numbering of levels from Table XLII. The capturing state is denoted by (*c*). All transitions are assumed to result from $S^{32}(n,\gamma)S^{33}$ reaction.

V. $S^{32}(d,p)S^{33}$ $Q_m=6.421$

By magnetic analysis at $E_d=1.8$, 2.9, and 6.5 Mev, the ground-state Q value$=6.422\pm0.011$ Mev (St 51), 6.408 ± 0.020 Mev (Le 56b), and 6.413 ± 0.006 Mev (Pa 57), respectively. Sixty-seven levels in S^{33} found recently by high-resolution magnetic analysis, Pa 57, Table XLIII, together with l_n values from angular distribution measurements (Ho 53a, Ho 53d). Levels in the $E_x=6.54$–7.2-Mev region not indicated in Fig. 25. Also, En 54a.

VI. $P^{33}(\beta^-)S^{33}$ See P^{33}

VII. $Cl^{33}(\beta^+)S^{33}$ See Cl^{33}

VIII. $S^{34}(\gamma,n)S^{33}$ $Q_m=-11.419$

Threshold measured as 10.85 ± 0.20 Mev (Sh 51a). Cross section, En 54a.

IX. $Cl^{35}(d,\alpha)S^{33}$ $Q_m=8.278$

From magnetic analysis, at $\theta=90°$, ground-state $Q=8.277\pm0.010$ Mev, and levels in S^{33} are observed at

TABLE XLII. Resonances in the sulfur total neutron cross section (Pe 50, Hu 55c).

E_n (kev)	E_x (S^{33}) (Mev)	J, π	l_n	Γ_n (kev)
111±2	8.754	1/2+	0	18±3
203±2	8.843		≥1	<2
274±2	8.912		≥1	<3
290±2	8.927		≥1	<3
375±3	9.010	1/2+	0	12±2
585±3	9.213	3/2−	1	1.4±0.5
700±4	9.325	1/2+	0	14±3

TABLE XLIII. Levels in S^{33} from $S^{32}(d,p)S^{33}$.

Level	E_x(S^{33}) in Mev [a]	l_n [b]	Level	E_x(S^{33}) in Mev [a]
(0)	...	2	(34)	5.888
(1)	0.839	0	(35)	5.915
(2)	1.965		(36)	5.982
(3)	2.314		(37)	6.067
(4)	2.869	} 3	(38)	(6.079)
(5)	2.936	} 3	(39)	6.101
(6)	2.971	} 3	(40)	6.131
(7)	3.222	1	(41)	6.234
(8)	3.832		(42)	6.261
(9)	3.935		(43)	6.310
(10)	4.049		(44)	6.326
(11)	4.095		(45)	6.360
(12)	4.145		(46)	6.372
(13)	4.211	1	(47)	6.416
(14)	4.377		(48)	6.427
(15)	4.425		(49)	6.487
(16)	4.732		(50)	6.513
(17)	4.747		(51)	6.526
(18)	4.869	} 1	(52)	6.559
(19)	4.919	} 1	(53)	6.616
(20)	4.941	} 1	(54)	6.675
(21)	5.177		(55)	6.689
(22)	5.210		(56)	6.710
(23)	5.272		(57)	6.720
(24)	5.287		(58)	6.788
(25)	5.340		(59)	6.892
(26)	5.351		(60)	6.903
(27)	5.399		(61)	6.965
(28)	5.479		(62)	6.999
(29)	5.597		(63)	7.017
(30)	5.613		(64)	7.037
(31)	5.622		(65)	7.133
(32)	5.711	1	(66)	7.164
(33)	5.864		(67)	7.183

[a] (Pa 57). Preliminary results obtained at $E_d=6.0$ and 6.5 Mev, $\theta=90°$ and 130°. Errors of 6 kev are assigned to all excitation energies, except to that of level (1) which has a 5-kev error. The proton groups leading to levels (5), (7), (32), (34), (48), (54), (55), and (67) are exceptionally intense.
[b] (Ho 53a, Ho 53d). Angular distribution measurements at $E_d=8.0$ Mev, using Al absorption for separation of proton groups.

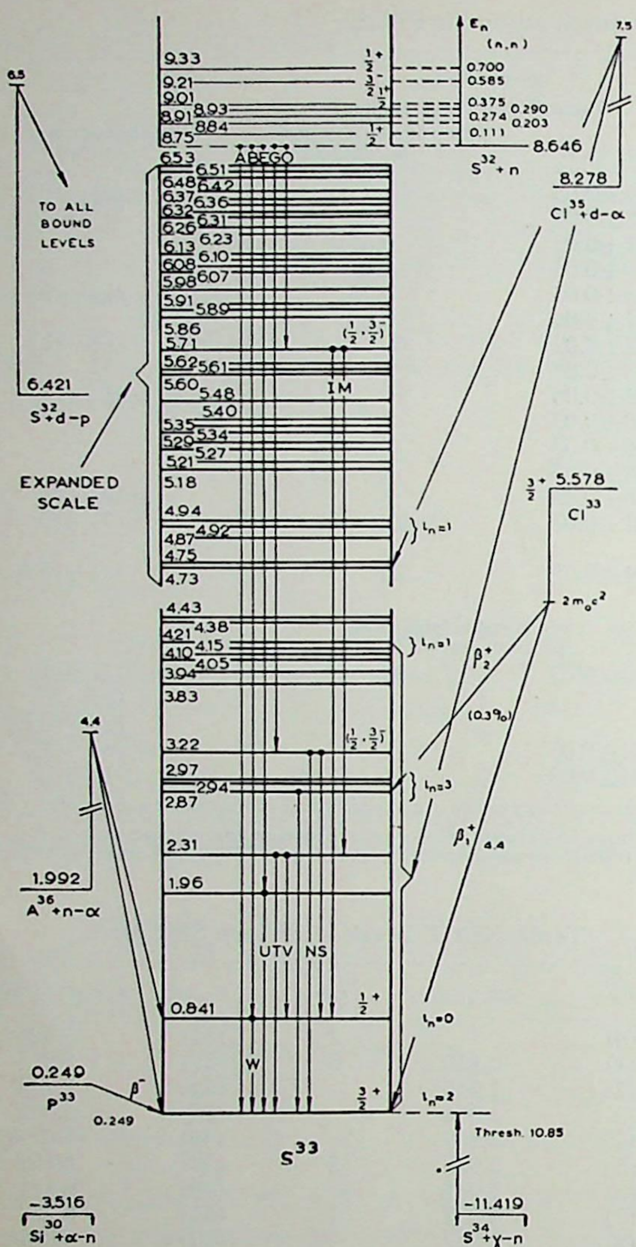


FIG. 25.

0.844±0.006, 1.966±0.007, 2.312, 2.869, 2.938, 2.969, 3.227, all ±0.008, 3.840, 3.947, 4.060, 4.105, 4.159, 4.224, all ±0.009, and 4.749±0.010 Mev. Barium-chloride targets containing natural Cl were bombarded with deuterons of several energies between 3.0 and 7.5 Mev (Pa 55c). A recent re-evaluation has shown that reported 3.365 Mev level does not exist. Alpha-particle groups leading to 4.38- and 4.43-Mev levels, observed from $S^{32}(d,p)S^{33}$ reaction, may have been present, but were weak or obscured by contaminant groups (Pa 57).

X. $A^{36}(n,\alpha)S^{33}$ $Q_m=1.922$

Ionization-chamber measurements with pulse-height analysis at $E_n=2.15$–4.40 Mev yield 2.0±0.1 Mev for ground-state Q value. A level at 1.1±0.1 Mev also observed (To 53). Also, En 54a.

Cl^{33}

(Fig. 26)

I. $Cl^{33}(\beta^+)S^{33}$ $Q_m=5.578$

Half-life measurements yield 2.8 sec (Ho 40) and 2.8 sec (Sc 48). Other determinations, 2.4±0.2 sec (Wh 41) and 1.8±0.1 sec (Bo 51), in doubt as possibly a mixture of Cl^{33} and Cl^{34} has been studied (St 53).

Possible confusion with short-lived Cl^{34} also sheds doubt on measurements of beta-spectrum end point: 4.13±0.07 Mev with a cloud chamber (Wh 41), and 4.43±0.13 Mev (Bo 51) and 4.2±0.2 Mev (Na 53) with a scintillation spectrometer. Log $ft=3.7$, approximately.

A γ ray of 2.85 Mev with an intensity of 0.3% per disintegration observed (Me 54).

II. $S^{32}(p,\gamma)Cl^{33}$ $Q_m=2.285$

In $E_p=200$–800 kev region, one resonance found at $E_p=594\pm5$ kev. Three γ rays with following energies and relative intensities observed:

γ_1 2.862±0.015 Mev (0.51),
γ_2 2.053±0.015 Mev (0.49),
γ_3 0.806±0.004 Mev (0.50).

Last γ ray is coincident with γ_2 but not with γ_1. Gamma-ray angular distribution measurements are only consistent with a $J=3/2^+$ assignment to the resonance and confirm assignments of $3/2^+$ to the ground state and $1/2^+$ to the 0.806 Mev level. Both γ_1 and γ_2 are of mixed M1+E2 character with mixing ratios of $x^2=0.10\pm0.01$ and 0.005±0.003, respectively (Va 56f).

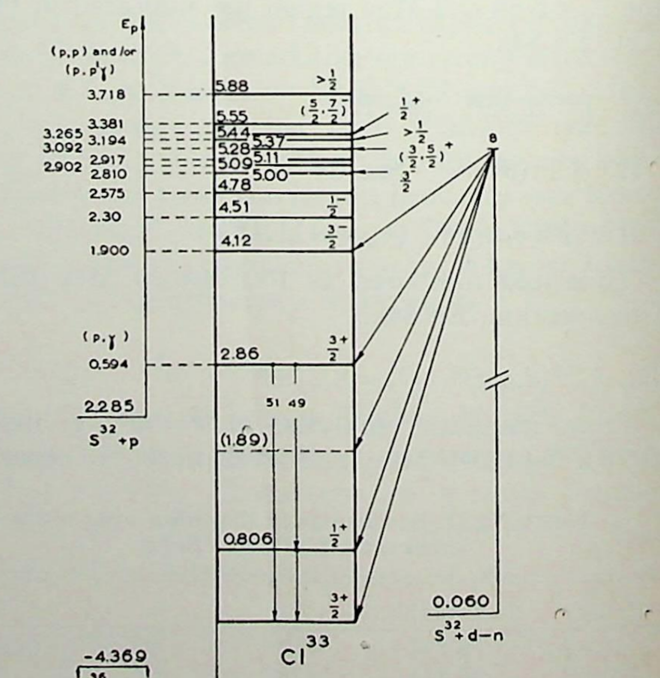


FIG. 26.

TABLE XLIV. Resonances for elastic proton scattering on S^{32}, and corresponding Cl^{33} levels (Ol 56).

E_p (Mev)	E_x (Cl^{33}) (Mev)	Γ (kev)	J, π
1.900	4.122	8.5	3/2−
2.30	4.51	52	1/2−
2.575	4.775	⩽5	
2.810	5.002	∼6	3/2−
2.902	5.092	⩽5	
2.917	5.107	⩽5	
3.092	5.275	⩽2	(3/2, 5/2)+
3.194	5.373	⩽2	>1/2
3.265	5.442	32	1/2+
3.381	5.554	⩽2	(5/2, 7/2)−
3.718	5.880	⩽2	>1/2

III. (a) $S^{32}(p,p)S^{32}$ $E_b=2.285$
(b) $S^{32}(p,p'\gamma)S^{32}$ $E_b=2.285$ $E_x=2.24$

Resonances observed in elastic scattering cross section at five different angles in Table XLIV. Essentially same results for $E_p=1.90$- and 2.30-Mev resonances also given, Fe 53. Two more very broad ($\Gamma\sim300$ kev) resonances probably present at about $E_p=3.5$ Mev (Ol 56). States at $E_p=2.81$, 2.90, 2.92, 3.09, 3.19, 3.38, and 3.72 Mev also resonant for 2.24-Mev γ rays from inelastic scattering, Le 56.

IV. $S^{32}(d,n)Cl^{33}$ $Q_m=0.060$

Several neutron groups observed with nuclear emulsions at $E_d=8$ Mev. Ground-state Q value is 0.25 Mev, and levels found at $E_x=0.76\pm0.07$, (1.89), 2.84 ± 0.06, and 4.22 ± 0.08 Mev. Several more closely spaced levels found above 4.22 Mev. From angular distribution measurements and Butler analysis, Cl^{33} ground state can be characterized as $J=(3/2,\ 5/2)^+$ ($l_p=2$), level at 0.76 Mev as $J=1/2^+$ ($l_p=0$), while levels at 2.84 and 4.22 Mev check best with $J=(1/2,\ 3/2)^-$ ($l_p=1$) (Mi 53).

V. $S^{33}(p,n)Cl^{33}$ $Q_m=-6.360$

Observed, En 54a.

VI. $A^{36}(p,\alpha)Cl^{33}$ $Q_m=-4.369$

Not observed.

P^{34}

(not illustrated)

I. $P^{34}(\beta^-)S^{34}$ $Q_m=5.2$

Best measurement of half-life yields 12.40 ± 0.12 sec (Bl 46, also Co 40, Hu 45).

The β^- decay proceeds predominantly to ground state and first excited state of S^{34} (Bl 46); in addition, the γ spectrum reveals weak branching to a 4.0-Mev level (Mo 56d). End-point of high-energy β^- branch is 5.1 ±0.2 Mev, that of second branch, 3.2 ± 0.2 Mev, as measured by Al absorption (Bl 46). Branching ratios and log ft in Table XLV.

TABLE XLV. The $P^{34}(\beta^-)S^{34}$ decay.

E_x (S^{34}) (Mev)	0	2.10	3.22	4.0 [b]
E_{β^-} (Mev)	5.1 ± 0.2 [a]	3.2 ± 0.2 [a]		
Branching ratio (%)	75 [a]	25 [a]	...	0.2 [b]
log ft	5.1 [a]	4.7 [a]	>5.6 [b]	4.9 [b]

[a] Bl 46. [b] Mo 56d.

Spin and parity of P^{34} very probably 1^+ (see S^{34}, general remarks).

Discussion of β^- transition probabilities, De 53b.

II. $P^{33}(n,\gamma)P^{34}$ $Q_m=6.5$

Perhaps observed, En 54a.

III. $S^{34}(n,p)P^{34}$ $Q_m=-4.4$

Cross section, En 54a.

IV. $S^{36}(d,\alpha)P^{34}$ $Q_m=4.8$

Not observed.

V. $Cl^{37}(n,\alpha)P^{34}$ $Q_m=-1.4$

Cross section, En 54a.

S^{34}

(Fig. 27)

I. $P^{31}(\alpha,p\gamma)S^{34}$ $Q_m=0.611$

Investigated with 8.1-Mev cyclotron α particles and with scintillation counters for energy and coincidence

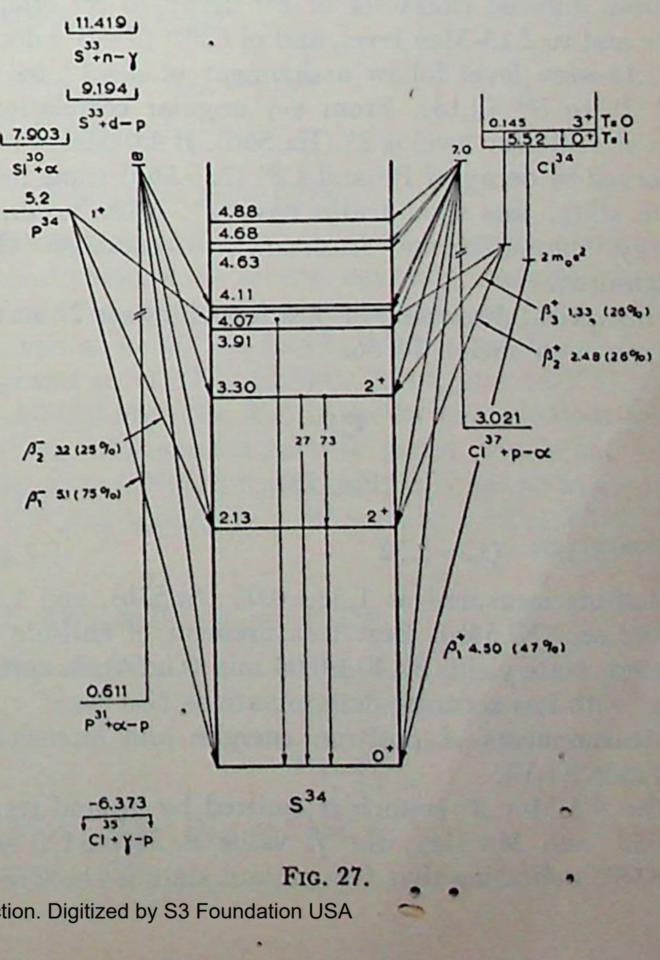


FIG. 27.

measurements of protons and γ quanta. Ground-state Q value $=0.7\pm0.1$ Mev; levels found at 2.13 ± 0.02, 3.33 ± 0.05, 4.32 ± 0.08, and 4.8 ± 0.1 Mev. Branching of γ decay of 3.30-Mev level is $(73\pm4)\%$ to 2.13-Mev level and $(27\pm4)\%$ to ground state (St 56b).

II. $S^{33}(n,\gamma)S^{34}$ $Q_m=11.419$

Not observed.

III. $S^{33}(d,p)S^{34}$ $Q_m=9.194$

Reaction reported (En 54a). Not confirmed (Pa 57).

IV. $P^{34}(\beta^-)S^{34}$ See P^{34}

V. $Cl^{34}(\beta^+)S^{34}$ See Cl^{34}

VI. $Cl^{35}(\gamma,p)S^{34}$ $Q_m=-6.373$

Not observed.

VII. $Cl^{37}(p,\alpha)S^{34}$ $Q_m=3.021$

Ground-state Q value measured by magnetic analysis as 3.015 ± 0.015 (Al 55a), 3.028 ± 0.006 (Va 56, Va 57d), and 3.026 ± 0.008 Mev (En 56c).

Also by magnetic analysis, at E_p between 1.8 and 4.0 Mev, a level was found at 2.129 ± 0.014 Mev (Va 56), while at $E_p=7.0$ Mev, levels found at 2.127, 3.302, 3.915, 4.073, 4.114, 4.621, 4.685, and 4.876, all ±0.008 Mev (En 56c).

GENERAL REMARKS

From allowed character of P^{34} decay to S^{34} ground state and to 2.13-Mev level, and of Cl^{34m} $(J=3^+)$ decay to 2.13-Mev level follow assignment of $J=1^+$ to P^{34} and 2^+ to S^{34} (2.13). From γ-γ angular correlations, spin of 3.30-Mev level is 2^+ (Ha 56d). If 4.0-Mev γ rays observed in decay of P^{34} and Cl^{34} (Mo 56d) come from same state, this should also have 2^+. More accurate energy measurements necessary to establish this assignment.

Theoretical discussion of position of lowest 2^+ states in even-even nuclei, Th 56.

Cl^{34}

(Fig. 28)

I. $Cl^{34}(\beta^+)S^{34}$ $Q_m=5.52$

Half-life measured as 1.58 ± 0.05 (St 53b), and 1.53 ± 0.02 sec (Kl 54b). Best measurement of half-life of isomeric state yields 32.40 ± 0.04 min (Gr 56), in agreement with less accurate determinations (En 54a).

Measurements of positron energies and intensities in Table XLVI.

The 4.5-Mev β^+ branch is emitted by ground state (Ar 53, also Ma 54a); the ft value is 3160 ± 120 sec (Gr 56), indicating that Cl^{34} ground state is the $T=1$, $J=0^+$ analog of S^{34} ground state (also Qu 56). Log ft for transitions from Cl^{34m} to S^{34} (2.13) and (3.30) are 6.1 and 4.9, respectively.

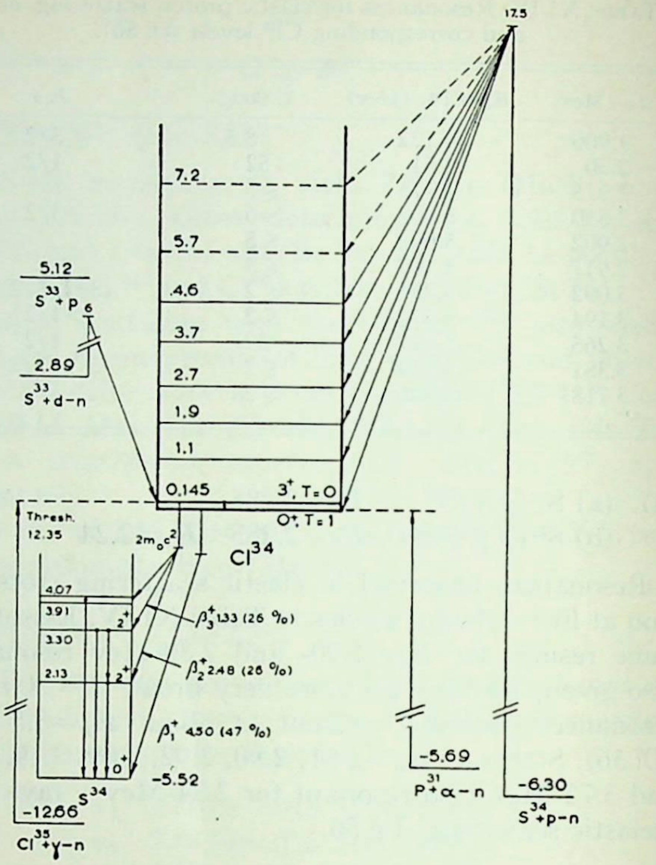


FIG. 28.

Energies of γ rays in decay of Cl^{34m} in Table XLVII. From conversion coefficient of 145-kev γ ray, which is $(14\pm4)\%$ (St 53), it follows that this is a M3 transition. Hence, isomeric state has $J=3^+$. Intensity of 4.0-Mev γ ray indicates log ft = about 5.4 for corresponding β^+ transition (Mo 56d). Angular correlation of 1.17- and 2.13-Mev γ rays indicates that both 2.13- and 3.30-Mev level have $J=2^+$. The 1.17-Mev γ ray has a 1.7% E2 admixture (Ha 56d). Discussion of spins and parities of S^{34} levels, under S^{34}, general remarks.

II. $P^{31}(\alpha,n)Cl^{34}$ $Q_m=-5.69$

Ground-state Q value, as derived from threshold measurements, is -5.7 ± 0.2 Mev. This corresponds to threshold for slow neutron production as well as for production of 1.58-sec ground-state activity (Qu 56).

TABLE XLVI. The β^+ decay of Cl^{34} and Cl^{34m}.

Reference: Method:	Ru 51 Magnetic spectrometer	Hu 54a Scintillation spectrometer	Gr 56 Magnetic spectrometer
End point β_1^+ (Mev)	4.45 ± 0.11 (46%)	4.45 ± 0.10	4.50 ± 0.03 (47%)
End point β_2^+ (Mev)	2.58 ± 0.26 (28%)		2.48 ± 0.07 (26%)
End point β_3^+ (Mev)	1.3 ± 0.2 (26%)		1.33 ± 0.10 (26%)

TABLE XLVII. Gamma rays in the decay of Cl^{34m}.

Reference: Method:	Ru 51 Magnetic spectrometer	Ti 51 Scintillation spectrometer	Mo 56d Scintillation spectrometer
E_{γ_1} (Mev)	0.145±0.003		
E_{γ_2} (Mev)		1.16±0.03	
E_{γ_3} (Mev)	2.13 ±0.12	2.10±0.03	
E_{γ_4} (Mev)	3.30 ±0.14	3.22±0.03	
E_{γ_5} (Mev)			4.0 (0.2%)

III. $S^{33}(p,\gamma)Cl^{34}$ $Q_m=5.12$

Not observed.

IV. $S^{33}(d,n)Cl^{34}$ $Q_m=2.89$

Observed, En 54a.

V. $S^{34}(p,n)Cl^{34}$ $Q_m=-6.30$

Studied with 17.5-Mev protons and detection of neutrons by nuclear emulsions. Highest Q value observed was -6.1 Mev; if this corresponds to Cl^{34} ground state, there are levels at 1.1, 1.9, 2.7, 3.7, 4.6, (5.7), and (7.2) Mev (Aj 55).

VI. $Cl^{35}(\gamma,n)Cl^{34}$ $Q_m=-12.66$

The Q value for transition to 32-min metastable state at 0.145 Mev is -12.35 ± 0.35 Mev (De 55). Cross section, En 54a, Go 54a.

GENERAL REMARKS

Positions of the lowest $T=0$ and $T=1$ states in Cl^{34} and other self-conjugated odd-odd nuclei, St 53b, St 53a, Pe 53, Mo 54, Wi 56a. Discussion of Cl^{34} ground-state configuration, Ku 53, De 53a. See Cl^{36}, General remarks, for discussion of relation between excitation energies in Cl^{34} and Cl^{36}.

S^{35}

(Fig. 29)

I. $S^{35}(\beta^-)Cl^{35}$ $Q_m=0.168$

Half-life averaged from four determinations (En 54a) is 87.3±1.1 days. Five magnetic or electrostatic spectrometer determinations (En 54a, Fe 54a) of β^- spectrum end point yield an average value of 168.1±0.5 kev. Spectrum has allowed shape, at least down to 2 kev (En 54a, Mo 54a, Di 55). Log $ft=5.0$. A spin $J=3/2$ for S^{35} found from microwave experiments (We 51). Computation of β^- matrix element based on jj coupling, Gr 56.

II. $S^{34}(n,\gamma)S^{35}$ $Q_m=6.988$

Cross section, En 54a, Hu 55c.

III. $S^{34}(d,p)S^{35}$ $Q_m=4.763$

From bombardment of natural sulfur targets at $E_d=6.5$ Mev and $\theta=90°$, weak proton groups observed by magnetic analysis corresponding to transitions to

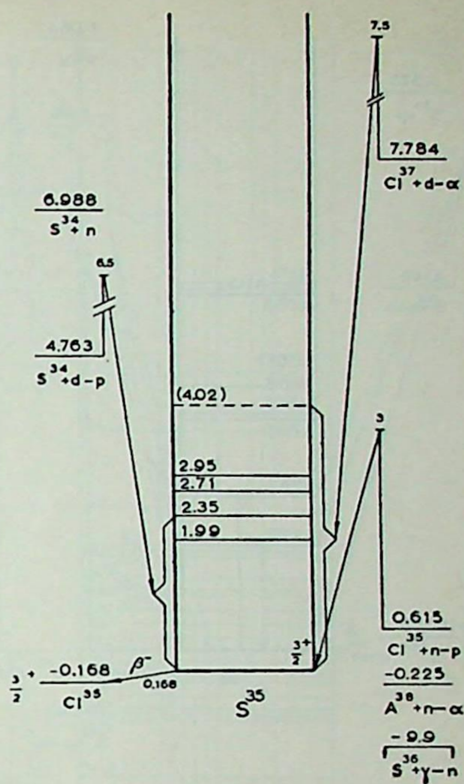


FIG. 29.

S^{35} ground state ($Q=4.76$ Mev) and to levels at 1.99 and 2.35 Mev (Pa 57).

IV. $Cl^{35}(n,p)S^{35}$ $Q_m=0.615$

Cross section, En 54a, Hu 55c, Be 55c. Q value determined from ionization-chamber pulse-height analysis both with thermal and with $D(d,n)$ neutrons as 0.52 ±0.04 Mev (Gi 44).

V. $S^{36}(\gamma,n)S^{35}$ $Q_m=-9.9$

Not observed.

VI. $Cl^{37}(d,\alpha)S^{35}$ $Q_m=7.784$

From bombardment of $BaCl_2$ targets containing natural chlorine, at several deuteron energies between 3.0 and 7.5 Mev, Q value determined by magnetic analysis at $\theta=90°$ as 7.783±0.012 Mev. Levels in S^{35} observed at (1.922), (2.348), 2.714, and (4.025) Mev, all ±0.010 Mev (Pa 55c). A recent re-evaluation shows that an additional α-particle group which had been assigned to $Cl^{35}(d,\alpha)$ reaction actually has to be ascribed to $Cl^{37}(d,\alpha)$ leading to a S^{35} level at 2.955±0.010 Mev (Pa 57).

VII. $A^{38}(n,\alpha)S^{35}$ $Q_m=-0.225$

Not observed.

Cl^{35}

(Fig. 30)

I. $S^{32}(\alpha,p)Cl^{35}$ $Q_m=-1.857$

Work with α particles from natural radioactive sources, En 54a.

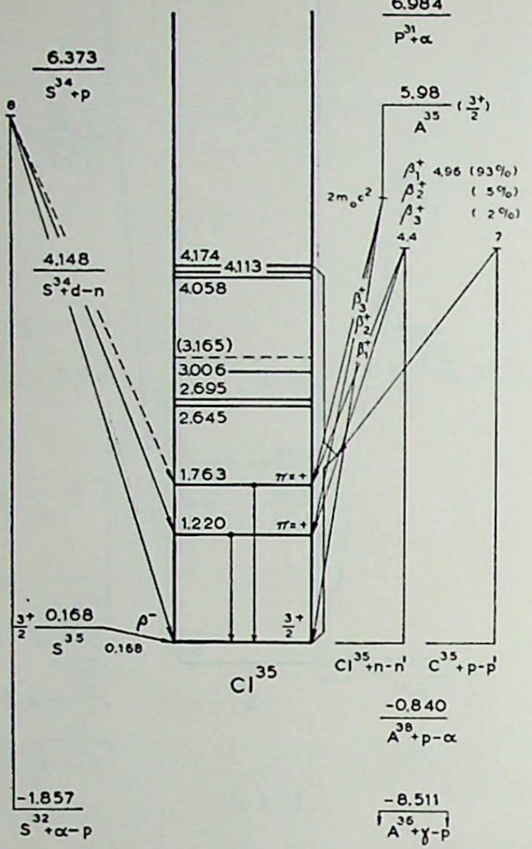


FIG. 30.

With 8-Mev cyclotron α particles and proton detection with a scintillation spectrometer, Q value determined as -2.3 Mev and levels in Cl^{35} observed at (0.7), 1.1, and (1.7) Mev (Pi 55).

II. $S^{34}(p,\gamma)Cl^{35}$ $Q_m=6.373$

With 1% energy resolution, resonances found in bombardment of sulfur with protons of 0.9 Mev $<E_p<1.9$ Mev at $E_p=1.37$, 1.61, 1.69, 1.8, and 1.86 Mev. As target shows no radioactivity these resonances are attributed to S^{34}, but, in view of short half-life of Cl^{33}, capture in S^{32} should not be excluded (Ha 51).

III. $S^{34}(d,n)Cl^{35}$ $Q_m=4.148$

Not observed.

TABLE XLVIII. Levels in Cl^{35} found from $Cl^{35}(p,p')Cl^{35}$ by magnetic analysis.

Reference:	Va 56	Sc 56e	En 56c
E_p (Mev):	3.6–3.8	4.6–5.6	7.0
Angle of observation:	90°		90° and 130°
Target:	natural Cl	enriched Cl	natural Cl
E_x (Cl^{35}) in Mev	1.219±0.005	1.220±0.005	1.221
	1.760±0.004	1.766±0.006	1.763
			2.645
			2.695
			3.006
			(3.165)
			4.058
			4.113
			4.174
			all ±0.005

IV. $S^{35}(\beta^-)Cl^{35}$ See S^{35}

V. $Cl^{35}(n,n'\gamma)Cl^{35}$

At $E_n=4.4$ Mev, γ rays of 1.23±0.01, and 1.77±0.01 Mev observed with a scintillation spectrometer, (Si 56).

VI. $Cl^{35}(p,p')Cl^{35}$

Levels in Cl^{35} given in Table XLVIII.

VII. $A^{35}(\beta^+)Cl^{35}$ See A^{35}

VIII. $A^{36}(\gamma,p)Cl^{35}$ $Q_m=-8.511$

Not observed.

IX. $A^{38}(p,\alpha)Cl^{35}$ $Q_m=-0.840$

Not observed.

A^{35}

(not illustrated)

I. $A^{35}(\beta^+)Cl^{35}$ $Q_m=5.98$

Determinations of half-life yield 2.2±0.2 sec (Wh 41), 1.88±0.04 sec (El 41), 1.84 sec (Sc 48), and 1.83±0.03 sec (Ki 56), giving average value of 1.85±0.03 sec.

The β^+ end point determined by magnetic spectrometer as 4.96±0.04 Mev. Weak γ rays of 1.19±0.04 and 1.73±0.04 Mev with intensities of 5±2 and 2±1%, respectively, observed by scintillation spectrometer. The β^+ transitions to Cl^{35} ground state, the first, and the second level have log $ft=3.8$, 4.7, and 4.5, respectively (Ki 56).

II. $S^{32}(\alpha,n)A^{35}$ $Q_m=-8.62$

Observed, En 54a.

III. $Cl^{35}(p,n)A^{35}$ $Q_m=-6.76$

Observed, En 54a, Ki 56.

IV. $A^{36}(\gamma,n)A^{35}$ $Q_m=-15.27$

Not observed.

S^{36}

(not illustrated)

I. $Cl^{36}(EC)S^{36}$ See Cl^{36}

II. $Cl^{37}(\gamma,p)S^{36}$ $Q_m=-8.4$

Not observed.

III. $A^{40}(\gamma,\alpha)S^{36}$ $Q_m=-6.8$

Alpha particles perhaps observed as background of large pulses from an argon-filled ionization chamber bombarded by 17.6-Mev γ rays (Wi 51).

Cl^{36}

(Fig. 31)

I. (a) $Cl^{36}(\beta^-)A^{36}$ $Q_m=0.714$
(b) $Cl^{36}(EC)S^{36}$ $Q_m=1.2$

Half-life measured by absolute β^- counting, using samples of known Cl^{36} content, as $(4.4\pm0.5)\times10^5$ yr (Wu 49) and $(3.08\pm0.03)\times10^5$ yr (Ba 55a).

End point of β^- spectrum is 0.714 ± 0.005 Mev (Fe 52, Wu 49). Shape of beta spectrum, investigated extensively with thin sources, is consistent with a $\Delta J=2$, no, transition (Wu 50, Fu 51, Fe 52, Jo 56). Log $ft=13.4$. Number of positrons is less than 10^{-4} times number of electrons (Wu 49, Jo 49), and no γ rays present with energy $E_\gamma>20$ kev and intensity of more than 5% (Wu 49).

Observation of K quanta in a proportional counter shows that Cl^{36} also decays by K capture to S^{36} with $K/\beta^-=(1.7\pm0.1)\%$ (Dr 55).

A spin $J=2$ obtained for Cl^{36} by microwave experiments (To 49, Jo 51, Gi 52a, Aa 55).

Theoretical discussions of Cl^{36} ground-state configuration, Ku 53, De 53a, Hi 54, Sc 54c. Discussion of log ft value, Fo 54, Jo 56.

II. $S^{33}(\alpha,p)Cl^{36}$ $Q_m=-1.923$

Not observed.

III. $Cl^{35}(n,\gamma)Cl^{36}$ $Q_m=8.580$

From thermal-neutron activation cross sections and abundances of Cl^{35} and Cl^{37}, it follows that probably less than 1% of thermal-neutron captures in natural chlorine occur in Cl^{37} (Hu 55e).

Gamma rays observed from capture of thermal neutrons in natural chlorine in Table XLIX. Apparently, many simple two-step cascades occur from capturing state through one Cl^{36} excited state to ground state. Cascades involving 0.790-, 1.16-, and 1.60-Mev levels confirmed by coincidence measurements with two scintillation spectrometers. A 735-kev γ ray (not in Table XLIX) observed in coincidence with line B can only be explained as transition to unknown 7.84-Mev level (Re 54).

IV. $Cl^{35}(n,n)Cl^{35}$ $E_b=8.580$

Cross section and resonances, Hu 55c, Br 56c. With enriched material, resonance observed at $E_n=26.5$ kev with $\Gamma=0.34$ kev (To 55).

V. $Cl^{35}(d,p)Cl^{36}$ $Q_m=6.355$

By magnetic analysis of (d,p) reactions with natural chlorine targets, at $\theta=90°$ and E_d between 3.0 and 7.5 Mev, ground-state $Q=6.354\pm0.008$ Mev (Pa 55c). Levels observed in Cl^{36} in Table L. From low-resolution work at $E_d=3.5$ Mev, additional levels found at 4.42, 5.08, 5.57, and 5.82 Mev (En 51b). Corresponding proton groups might also result from $Cl^{37}(d,p)Cl^{38}$ reaction. Also En 54a.

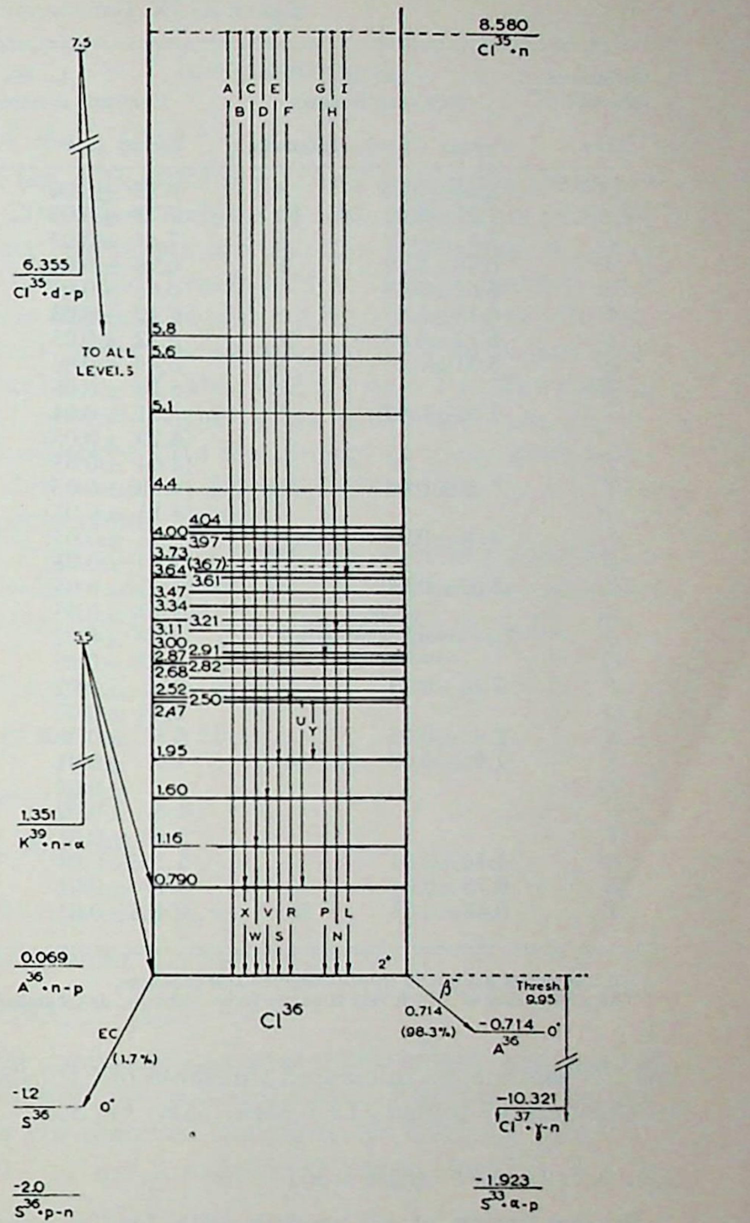


FIG. 31.

Angular distribution of ground-state proton group measured with nuclear emulsions at $E_d=6.90$ Mev and 7.8 Mev. Butler analysis yields $l_n=2$ in both cases, with less than 4% admixture of $l_n=0$ (Ki 52a, Ki 53b).

VI. $S^{36}(p,n)Cl^{36}$ $Q_m=-2.0$

Not observed.

VII. $A^{36}(n,p)Cl^{36}$ $Q_m=0.069$

Not observed.

VIII. $Cl^{37}(\gamma,n)Cl^{36}$ $Q_m=-10.321$

Photoneutron threshold of natural chlorine is 9.95 ±0.20 Mev (Sh 51a). This value should be assigned to

TABLE XLIX. Gamma rays from thermal neutron capture in chlorine.

Reference: Method: Line	Ki 52 Pair spectrometer Energy (Mev)	Intensity [a]	Gr 55a Compton spectrometer Energy (Mev)	Intensity [a]	Ha 52b Scintillation spectrometer Energy (Mev)	Re 53 Two-crystal scintillation spectrometer Energy (Mev)	Intensity [a]	Probable transition in Cl^{36} [b]
A	8.56±0.03	3	8.55 ±0.04	2.8				(*c*)→(0)
B	7.77±0.03	10	7.78 ±0.03	7.8	7.7 ±0.2			(*c*)→(1)
C	7.42±0.03	8	7.41 ±0.03	14				(*c*)→(2)
D	6.98±0.03	1	6.96 ±0.04	1.9				(*c*)→(3)
E	6.62±0.06	4	6.64 ±0.04	14.4				(*c*)→(4)
F	6.12±0.03	6	6.12 ±0.03	21.4	6.2 ±0.2			(*c*)→(5)
G	5.72±0.03	2	5.72 ±0.03	5.6				(*c*)→(10)
H	5.51±0.03	1	5.49 ±0.04	2				(*c*)→(13)
H′			(5.28 ±0.05)	1.6				
I	5.01±0.03	4	5.01 ±0.04	6		4.67±0.10		(*c*)→(17)
I′			(4.79 ±0.05)	1.9				
I″			(4.64 ±0.05)	2.3				
J	4.46±0.04	2	(4.50 ±0.05)	2.2				
J′			(4.15 ±0.05)	2.3				
K	4.06±0.04	3	(4.05 ±0.05)	2.1		3.71±0.10	23	
K′			(3.90 ±0.05)	1.8				
L	3.62±0.05	2	(3.63 ±0.05)	2.9				(17)→(0)
M	Br 56e		3.40 ±0.05	3.6				
N	Two crystal scintillation		3.09 ±0.02	} 8.0				
O	spectrometer		3.02 ±0.03					(13)→(0)
P	2.84±0.03	6	2.87 ±0.02	9.5		2.68±0.10	10	(10)→(0)
Q			2.68 ±0.02	2.0				
R	2.42±0.05	1	2.51 ±0.03	1.0		2.03±0.10		(5)→(0)
S	1.96±0.03	34	1.97 ±0.01	29	2.00 ±0.025	1.77±0.10	19	(4)→(0)
T			1.72 ±0.02	1				
U			1.67 ±0.02	1.0				(5)→(1)
V			1.60 ±0.01	2.4	1.59 ±0.013			(3)→(0)
W	1.14±0.03	8	1.165±0.01	36	1.15 ±0.01	1.12±0.10	29	(2)→(0)
X	0.75±0.03	10	0.77 ±0.01	23	0.784±0.01	0.70±0.10	35	(1)→(0)
Y	0.48±0.03	10	0.485±0.01	26				(5)→(4)

[a] All intensities are given in photons per 100 captures.
[b] The numbering of Cl^{36} levels is taken from Table L; (*c*) stands for the capturing state.

the Cl^{37} isotope, as calculated threshold of $Cl^{35}(\gamma,n)Cl^{34}$ is considerably higher (12.7 Mev). Also Go 54a.

IX. $K^{39}(n,\alpha)Cl^{36}$ $Q_m=1.351$

By irradiation of KI crystals with D(*d*,*n*) neutrons and pulse-height analysis of scintillation spectrum, ground-state $Q=1.25\pm0.2$ Mev. Transition to a Cl^{36} level at 0.87±0.1 Mev also observed (Sc 56a).

GENERAL REMARKS

Pure *jj* coupling predicts a linear relation between excitation energies of $(d_{3/2}, d_{3/2}^{-1})$ levels in Cl^{36} and

TABLE L. Levels in Cl^{36} from $Cl^{35}(d,p)Cl^{36}$ (Pa 55c).

Level	E_x (Cl^{36}) (Mev)	Level	E_x (Cl^{36}) (Mev)
(1)	0.790±0.005	(13)	3.110±0.008
(2)	1.163±0.006	(14)	3.214±0.008
(3)	1.600±0.007	(15)	3.341±0.008
(4)	1.952±0.007	(16)	3.474±0.008
(5)	2.473±0.007	(17)	3.606±0.008
(6)	2.498±0.007	(18)	3.644±0.008
(7)	2.523±0.007	(19)	(3.673±0.008)
(8)	2.684±0.007	(20)	3.732±0.008
(9)	2.820±0.007	(21)	3.970±0.008
(10)	2.872±0.007	(22)	4.003±0.008
(11)	2.905±0.007	(23)	4.043±0.008
(12)	3.004±0.007		

those of $(d_{3/2}, d_{3/2})$ levels in Cl^{34} (Pa 56a). Most natural way to obtain in Cl^{34} a 0^+ ground state, a 3^+ 142-kev level, and a 2^+ level at about 2.13 Mev (as in S^{34}), would be to assign 2^+, 3^+, 1^+, and 0^+ to Cl^{36} ground state and to levels at 0.790, 1.60, and 3.34 Mev, respectively. This entails 1^+ and 2^+ levels in Cl^{34} at 0.51 and 1.80 Mev, respectively.

A^{36}

(Fig. 32)

I. $S^{33}(\alpha,n)A^{36}$ $Q_m=-1.922$

Not observed.

II. $Cl^{35}(p,\gamma)A^{36}$ $Q_m=8.511$

Eighty-six resonances in γ-ray yield from targets containing natural chlorine found at proton energies between 500 and 2150 kev. From targets enriched in Cl^{35}, resonances observed at 858, 888, 1102, 1258, 1484, and 1510 kev (Br 51).

Gamma-ray spectra measured at the 858-, 1484-, and 1510-kev resonances, and at an additional resonance at $E_p=1580$ kev which also has to be assigned to Cl^{35}.

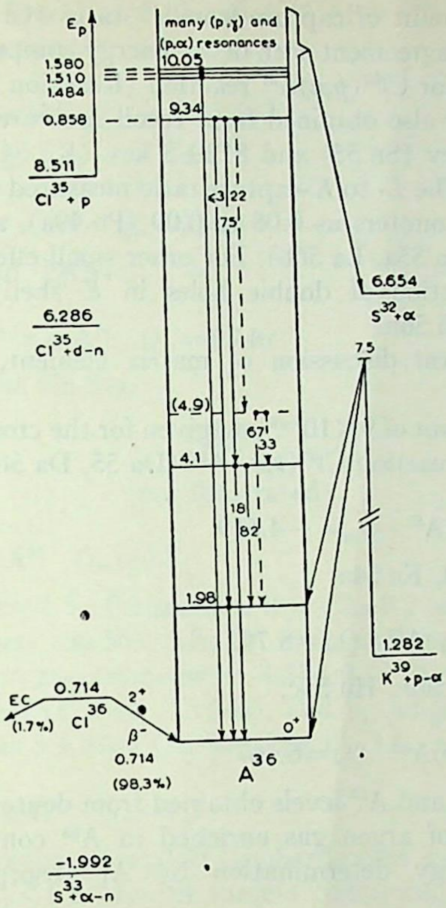


FIG. 32.

These have $(2J+1)\Gamma_\gamma=9.6$, 8.0, 32, and 22 ev, respectively. The 858-kev resonance decays through levels at 1.95, 4.1, and (4.9) Mev. Branching ratios in Fig. 32. Coincidence measurements confirmed 5.1—4.1- and 5.1—2.15—1.95-Mev cascades. Other three resonances decay directly to A^{36} ground state. Measured γ-ray energies lead to $Q=8.4\pm0.15$ Mev (To 57).

III. $Cl^{35}(p,\alpha)S^{32}$ $E_b=8.511$ $Q_m=1.857$

Eighteen resonances observed in $E_p=1.30$–2.51-Mev region (Al 55a). For Q values, see S^{32}.

IV. $Cl^{35}(d,n)A^{36}$ $Q_m=6.286$

Not observed.

V. $Cl^{36}(\beta^-)A^{36}$ See Cl^{36}

VI. $K^{39}(p,\alpha)A^{36}$ $Q_m=1.282$

Magnetic analysis at $E_p=1.9$ Mev, ground state $Q=1.267\pm0.020$ Mev (Al 55a), while at E_p between 6.0 and 7.5 Mev, $Q=1.286\pm0.010$ Mev and a level in A^{36} at 1.977 ± 0.008 Mev have been observed (Sp 55, Sp 55b).

GENERAL REMARKS

Lowest $T=1$ level in A^{36} expected at 6.06 ± 0.06 Mev (Wi 56a).

S^{37}

(not illustrated)

I. $S^{37}(\beta^-)Cl^{37}$ $Q_m=4.7$

Half-life, 5.04 ± 0.02 min. The β^- spectrum, measured by Al absorption, consists of two components with the following end points and relative intensities: 4.3 ± 0.3 Mev (10%) and 1.6 ± 0.1 Mev (90%). Second component in coincidence with a 2.7 ± 0.2-Mev γ ray (Bl 46). Essentially same results obtained by scintillation spectrometer: $E_{\beta_1}\sim4.7$ Mev (10%), $E_{\beta_2}=1.6$ Mev (90%), $E_\gamma=3.1$ Mev. No other γ ray with more than 1% intensity observed (Mo 56). Energy of γ ray more accurately determined as 3.09 ± 0.03 Mev, also by scintillation spectrometer (St 56a).

The β^- ground-state transition is evidently forbidden [$\log ft=7.2$ and $\log (W_0^2-1)\ ft=9.2$] while transition to 3.1-Mev level in Cl^{37} is allowed ($\log ft=4.2$).

II. $S^{36}(n,\gamma)S^{37}$ $Q_m=4.4$

Cross section, Hu 55c.

III. $S^{36}(d,p)S^{37}$ $Q_m=2.2$

Not observed.

IV. $Cl^{37}(n,p)S^{37}$ $Q_m=-3.9$

Cross section, En 54a, Co 56.

V. $A^{40}(n,\alpha)S^{37}$ $Q_m=-2.3$

From ionization-chamber pulse-height analysis, at $E_n=14$ Mev, $Q=-2.5\pm0.1$ Mev. Four more α-particle groups are observed leading to S^{37} levels at 1.3 ± 0.05, 2.2 ± 0.1, 2.7 ± 0.1, and 3.5 ± 0.2 Mev (Be 55a).

Cl^{37}

(Fig. 33)

I. $Cl^{36}(n,\gamma)Cl^{37}$ $Q_m=10.321$

Cross section, Bo 55a.

II. $S^{37}(\beta^-)Cl^{37}$ See S^{37}

III. $Cl^{37}(p,p')Cl^{37}$

From bombardment of natural chlorine targets with 7.0-Mev protons and magnetic analysis at $\theta=90°$ and 130°, levels in Cl^{37} observed at 0.838, 1.728, 3.087, and 3.105 Mev, all ±0.005 Mev (En 56c). With enriched Cl^{37} targets, magnetic analysis at E_p between 4.6 and 5.6 Mev yields only one level below $E_x=2$ Mev, at $E_x=1.713\pm0.010$ Mev (Sc 56c).

IV. $A^{37}(EC)Cl^{37}$ See A^{37}

No other reactions leading to Cl^{37} have been observed. It can be reached in principle by $S^{34}(\alpha,p)Cl^{37}$

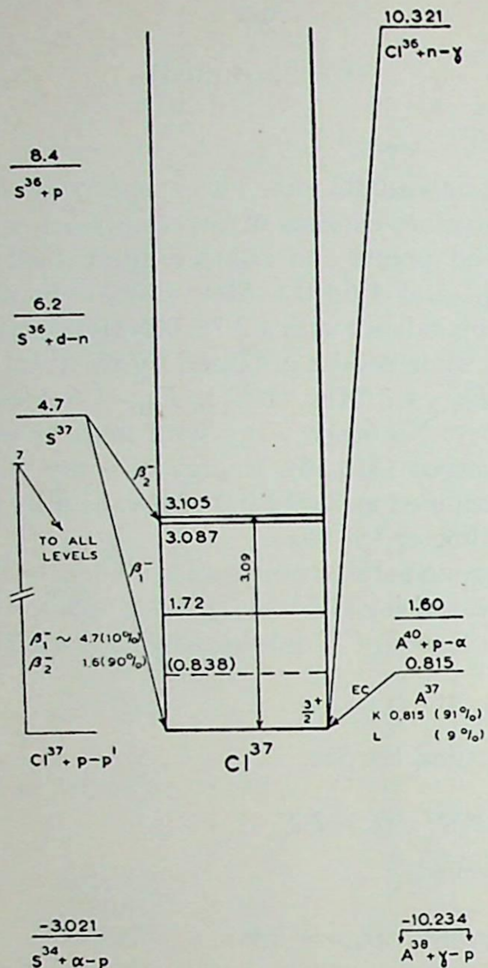


FIG. 33.

($Q_m=-3.021$), $S^{36}(p,\gamma)Cl^{37}$ ($Q_m=8.4$), $S^{36}(d,n)Cl^{37}$ ($Q_m=6.2$), $A^{38}(\gamma,p)Cl^{37}$ ($Q_m=-10.234$), and $A^{40}(p,\alpha)$ Cl^{37} ($Q_m=1.60$).

A^{37}

(Fig. 34)

I. $A^{37}(EC)Cl^{37}$ $Q_m=0.815$

Half-life averaged from two measurements (En 54a), 34.4±0.4 days. Log $ft=5.0$.

Shape and intensity of internal bremsstrahlung spectrum investigated by several authors (An 53, Em 54, Li 55b, Sa 56). Results are in agreement with theory taking account of capture from P states (Gl 56). End point is in agreement with decay energy computed from threshold for $Cl^{37}(p,n)A^{37}$ reaction (Reaction III). Decay energy also obtained from recoil measurements as 827±16 kev (Sn 55) and 812±8 kev (Ko 54, Wi 55a, Ru 55b). The L- to K-capture ratio measured with proportional counters as 0.08 to 0.09 (Po 49a), and 0.092 ±0.010 (La 55a, La 56b). For other small effects (e.g., the production of double holes in K shell), Mi 54, Wo 54a, Wi 56b.

TABLE LI. Levels in A^{37} from $A^{36}(d,p)A^{37}$.

Author: E_d (Mev):	Da 49 3.4		Zu 50 3.9	
	Q values (Mev)	A^{37} levels (Mev)	Q values (Mev)	A^{37} levels (Mev)
	6.59±0.03	...	6.49±0.08	...
	5.06±0.03	1.53	5.05±0.05	1.44
	4.92±0.05	1.67		
	4.32±0.05	2.27		
	4.03±0.03	2.56	3.93±0.05	2.56
	3.13	3.46	2.95±0.05	3.54
			2.09±0.07	4.40
			1.86±0.07	4.63
	1.58	5.01	1.42±0.03	5.07
			0.64±0.07	5.85

Theoretical discussion of matrix element, Go 56b, Gr 56c.

Upper limit of 9×10^{-46} cm² given for the cross section of inverse reaction $Cl^{37}(\bar{\nu},e^-)A^{37}$ (Da 55, Da 56b).

II. $S^{34}(\alpha,n)A^{37}$ $Q_m=-4.619$

Observed, En 54a.

III. $A^{36}(n,\gamma)A^{37}$ $Q_m=8.792$

Cross section, Hu 55c.

IV. $A^{36}(d,p)A^{37}$ $Q_m=6.567$

Q values and A^{37} levels obtained from deuteron bombardment of argon gas enriched in A^{36} content and proton-energy determination by Al absorption, in Table LI.

V. $Cl^{37}(p,n)A^{37}$ $Q_m=-1.598$

Threshold measured as $E_p=1.60$ Mev (Bl 51), 1640±4 kev (Ri 50), and 1641±2 kev (Sc 52a). By nuclear emulsion technique, neutron groups found with $Q=-1.58$ and -2.99 Mev corresponding to a A^{37} level at 1.41±0.05 Mev (St 52a). By the same method,

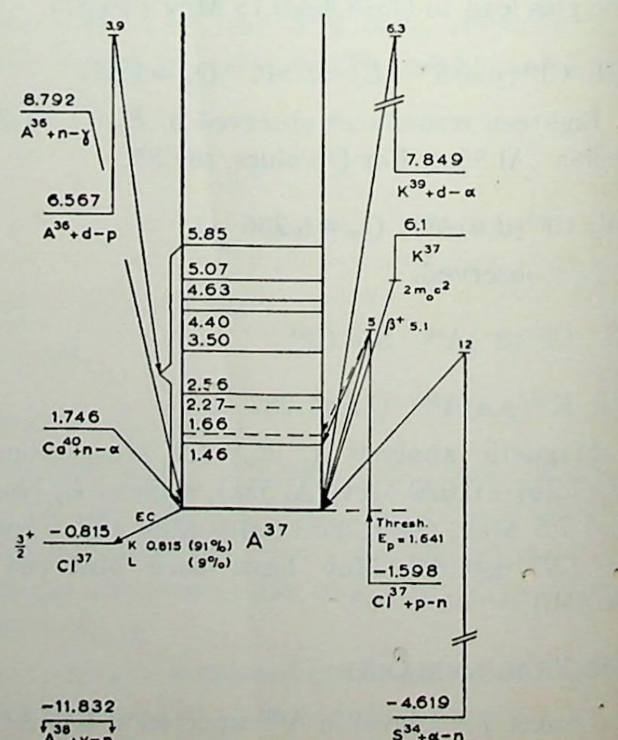


FIG. 34.

A^{37} levels found at 1.40 and 1.65 Mev (Gr 50b). See A^{38} for resonances.

VI. $A^{38}(\gamma,n)A^{37}$ $Q_m=-11.832$

Not observed.

VII. $K^{39}(d,\alpha)A^{37}$ $Q_m=7.849$

Observed, En 54a.

VIII. $Ca^{40}(n,\alpha)A^{37}$ $Q_m=1.746$

Observed, En 54a.

K^{37}

(not illustrated)

I. $K^{37}(\beta^+)A^{37}$ $Q_m=6.1$

Half-life=1.3±0.1 sec (La 48), 1.2±0.2 sec (Bo 51), and 1.2 sec (Su 56). The β^+ end point measured by scintillation spectrometer as 4.57±0.13 Mev (Bo 51), and 5.06±0.11 Mev (Ki 54b), and by magnetic spectrometer as 5.1 Mev (Su 56). Log $ft=3.6$.

II. $A^{36}(p,\gamma)K^{37}$ $Q_m=1.9$

No resonances found in bombardments with protons of $0.5<E_p<1.8$ Mev of targets containing separated A^{36} (Br 48a).

III. $A^{36}(d,n)K^{37}$ $Q_m=-0.3$

Not observed.

IV. $Ca^{40}(p,\alpha)K^{37}$ $Q_m=-5.2$

Observed, Su 56.

Cl^{38}

(Fig. 35)

I. $Cl^{38}(\beta^-)A^{38}$ $Q_m=4.910$

Best measurement of half-life yields 37.29±0.04 min (Co 50a), in good agreement with other less accurate measurements (En 54a, Ma 55b).

The β^- spectrum is complex and can be analyzed into three branches. Two γ rays observed. Energies and relative intensities determined by magnetic spectrometer in Table LII. Intensity of a potential 3.75-Mev cross-over γ ray is less than 0.03% per β^- particle (My 49). Shape of the high-energy β^- spectrum and its ft value [log $ft=7.5$ and log $(W_0^2-1)\ ft=9.5$] characteristic for once-forbidden transitions ($J=2$, yes) (La 50, Wu 50). The β^- transitions to 2.15-Mev level also forbidden (log $ft=6.9$) while those to 3.75-Mev level allowed (log $ft=4.9$). From data given above and from measured γ-γ angular correlation (Wa 41, Ki 42, St 50, Kr 54a), spins and parities can be assigned as shown in A^{38} level diagram.

TABLE LII. Beta decay of Cl^{38}.

Reference:	Wa 39	Cu 40	It 41	Ho 46	La 50
E_{β_1} (Mev)	4.99±0.06			5.2 (53%)	4.81±0.05 (53.4%)
E_{β_2} (Mev)				2.70 (11%)	2.77±0.05 (15.8%)
E_{β_3} (Mev)	1.08±0.06			1.19 (36%)	1.11±0.01 (30.8%)
E_{γ_1} (Mev)		2.15 (57%)	2.19±0.03 (53%)	2.15 (57%)	
E_{γ_2} (Mev)		1.65 (43%)	1.64±0.02 (47%)	1.60 (43%)	

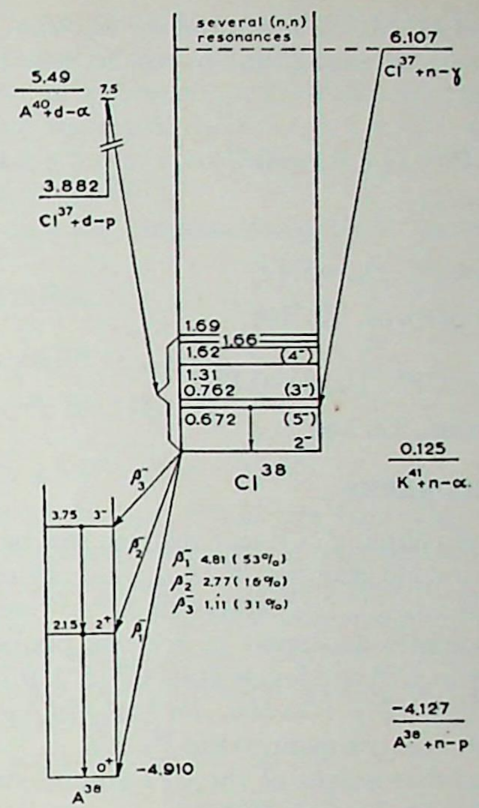


FIG. 35.

II. $Cl^{37}(n,\gamma)Cl^{38}$ $Q_m=6.107$

Cross section, Hu 55c.

From thermal neutron capture in concentrated Cl^{37}, an isomeric state found emitting γ rays of $E_\gamma=660\pm20$ kev with half-life of 1.0±0.2 sec (Sc 54).

III. $Cl^{37}(n,n)Cl^{37}$ $E_b=6.107$

Resonances in total neutron cross section observed with enriched material (neutron energies in kev, widths between brackets in kev): 8.8 (0.09), 26.5 (0.33), 47.4 (0.20), 55.5 (0.18), 65.0 (0.18), and 94.2 (0.80) (To 55).

IV. $Cl^{37}(d,p)Cl^{38}$ $Q_m=3.882$

Targets containing natural Cl bombarded with deuterons of several energies between 3.0 and 7.5 Mev. By magnetic analysis at $\theta=90°$, ground-state $Q=3.881\pm0.008$ Mev, and levels in Cl^{38} observed at 0.672±0.005, 0.762±0.005, 1.312±0.006, 1.626±0.007, 1.658±0.007, and 1.693±0.007 Mev. Intensity ratio of proton groups leading to ground state and first excited

state is 0.43±0.04, as measured at $E_d=5.6$ Mev, in agreement with spins of 2^- and 5^- respectively (Pa 55c). Also, En 54a.

V. $A^{38}(n,p)Cl^{38}$ $Q_m=-4.127$

Not observed.

VI. $A^{40}(d,\alpha)Cl^{38}$ $Q_m=5.49$

Perhaps observed, En 54a.

VII. $K^{41}(n,\alpha)Cl^{38}$ $Q_m=-0.125$

Cross section, En 54a.

GENERAL REMARKS

For pure *jj* coupling, a linear relation can be derived between excitation energies of $(d_{3/2}, f_{7/2})$ states in Cl^{38} and those of $(d_{3/2}^{-1}, f_{7/2})$ states in K^{40}. From known excitation energies and spins in K^{40} one predicts a 2^- ground state in Cl^{38} and levels at 0.702 (5^-), 0.751 (3^-), and 1.328 Mev (4^-) (Go 56a, Pa 56a), in very good agreement with experiment. Also, Pa 57a.

Theoretical discussions of the Cl^{38} ground-state configuration, Ku 53, De 53a, Hi 54.

A^{38}

(Fig. 36)

I. $Cl^{35}(\alpha,p)A^{38}$ $Q_m=0.840$

From bombardment of enriched Cl^{35} targets with 7.5-Mev cyclotron α particles, and proton energy determination by Al absorption, $Q=0.81\pm0.08$ Mev, and levels in A^{38} observed at 2.13±0.04 and 3.73±0.04 Mev (Kr 53).

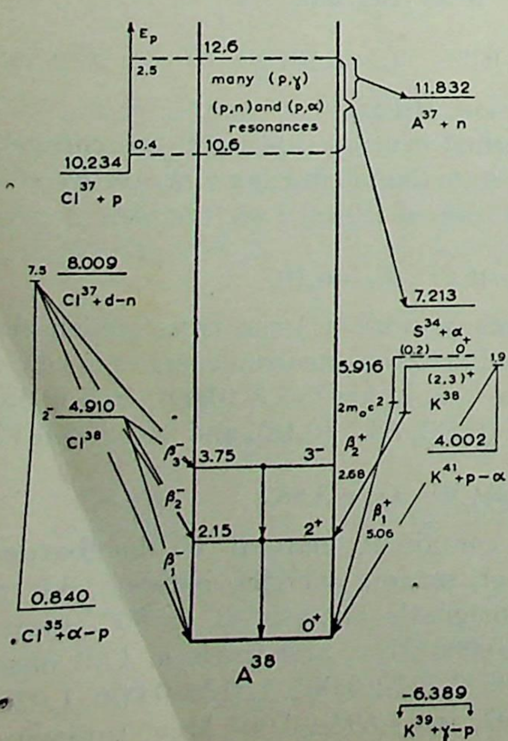


FIG. 36

Energies of γ rays measured by pair formation in a β spectrometer as $E_\gamma=1.7\pm0.2$, 2.7±0.3, and 3.9±0.4 Mev (Ma 37, Ma 41). Also, En 54a.

II. (a) $Cl^{37}(p,\gamma)A^{38}$ $E_b=10.234$
(b) $Cl^{37}(p,n)A^{37}$ $E_b=10.234$ $Q_m=-1.598$
(c) $Cl^{37}(p,\alpha)S^{34}$ $E_b=10.234$ $Q_m=3.021$

Many resonances observed in these reactions, En 54a, Al 55a, To 57.

Gamma-ray spectra measured at $E_p=1070$-, 1135-, 1533-, and 1725-kev resonances. First three have $(2J+1)\Gamma_\gamma=9$, 14, and 28 ev, respectively. Decay at 1070- and 1135-kev resonances proceeds mainly to 2.15- and 3.75-Mev levels in A^{38}, and to a new level at about 5 Mev, decaying to ground state and 2.15-Mev level. The 1533-kev resonance decays to 2.15- and 5-Mev levels, while 1725-kev resonance decays mainly to a new 6.7-Mev level, which is de-excited to 3.75-Mev level. The $Q=10.15\pm0.15$ Mev from γ-ray energies (To 57).

III. $Cl^{37}(d,n)A^{38}$ $Q_m=8.009$

Not observed.

IV. $Cl^{38}(\beta^-)A^{38}$ See Cl^{38}

V. $K^{38}(\beta^+)A^{38}$ See K^{38}

VI. $K^{39}(\gamma,p)A^{38}$ $Q_m=-6.389$

Not observed.

VII. $K^{41}(p,\alpha)A^{38}$ $Q_m=4.002$

Magnetic analysis at $E_p=1.9$ Mev and $\theta=90°$, gives $Q=4.002\pm0.015$ Mev (Al 55a).

GENERAL REMARKS

Theoretical discussions on low-lying excited states in A^{38}, Ta 54, Th 56.

K^{38}

(not illustrated)

I. (a) $K^{38}(\beta^+)A^{38}$ $Q_m=5.916$

Weighted average of all half-life measurements for which errors have been given (En 54a, Gr 56), 7.62 ±0.06 min.

The β^+ end point measured with lens spectrometer is 2.68±0.03 Mev. Spectrum has allowed shape. Log *ft* =5.0. Intensity of possible higher energy β^+ transitions smaller than 0.6% of main transition. No conversion lines observed (Gr 56).

The β^+ decay is followed by a γ ray with energy $E_\gamma=2.16\pm0.03$ Mev as measured by scintillation spectrometer (Ti 51).

Results given above limit K^{38} spin to $(2,3)^+$.

For *jj*-coupling calculation of β^+ matrix element, Gr 56c.

Also En 54a.

(b) $K^{38m}(\beta^+)A^{38}$

Half-life, 0.95 ± 0.03 sec (St 53b) and 0.935 ± 0.025 sec (Kl 54b).

The β^+ end point measured with scintillation spectrometer as 5.06 ± 0.11 Mev (Hu 54a). Possible confusion with K^{37}.

Measured end point would place isomeric state about 0.2 Mev above K^{38} ground state. The β^+ transition is superallowed ($\log ft=3.5$), corresponding to a $J=0^+$, $T=1$, assignment to isomeric state. Occurrence of isomerism requires a spin difference with the ground state of at least 3, which determines K^{38} ground state as 3^+. Theoretical discussions concerning the isobaric spin of levels in K^{38}, see Mo 54, Ta 54, Wi 56a.

II. $Cl^{35}(\alpha,n)K^{38}$ $Q_m=-5.859$

Observed, En 54a.

III. $A^{38}(p,n)K^{38}$ $Q_m=-6.699$

Not observed.

IV. $K^{39}(\gamma,n)K^{38}$ $Q_m=-13.088$

Threshold for production of 7.6-min activity, 13.2 ± 0.2 Mev (Mc 49), and 13.00 Mev (De 55). Cross section, En 54a.

V. $Ca^{40}(d,\alpha)K^{38}$ $Q_m=4.651$

Q value of transition to lowest $T=0$ state$=4.650\pm0.010$ Mev by magnetic analysis at $\theta=90°$ and at several deuteron energies up to 7 Mev. Level observed with $Q=4.20\pm0.01$ Mev (Br 56f). Observed excited state has also very probably $T=0$.

Cl^{39}

(not illustrated)

I. $Cl^{39}(\beta^-)A^{39}$ $Q_m=3.44$

Half-life, 55.5 ± 0.2 min (Ha 50a) and 56.5 min (Ru 52).

Beta decay is complex. By magnetic spectrometer, three branches observed with following end points and relative intensities: 3.45 ± 0.02 Mev ($7\pm2\%$), 2.18 Mev ($8\pm4\%$), and 1.91 ± 0.02 Mev ($85\pm6\%$). Shape of high-energy branch characteristic of unique, once-forbidden transition ($\Delta J=2$, yes). By scintillation spectrometer, three γ rays observed with following energies and relative intensities: $E_{\gamma_1}=0.246\pm0.003$ Mev (0.9 ± 0.1), $E_{\gamma_2}=1.266\pm0.010$ Mev (1), $E_{\gamma_3}=1.520\pm0.010$ Mev (0.85 ± 0.05). Upper limits of 2.1×10^{-3}, 1.6×10^{-3}, and 3.2×10^{-3} for conversion coefficients of γ_1, γ_2, and γ_3, respectively. Coincidences observed of γ_1 with γ_2, while γ_3 is not coincident with γ_1, nor with γ_2. The β^- spectrum, coincident with γ_1, γ_2, and γ_3 has an end point at 1.90 ± 0.05 Mev.

The β-γ coincidences delayed; half-life, $(9.5\pm0.5)\times10^{-10}$ sec.

Log ft for β^- transitions to A^{39} (0), (1.27), and (1.52) $=8.3$, 6.8, and 5.6, respectively.

All data in agreement with assignments of $3/2^+$ to Cl^{39} ground state, and of $7/2^-$, $(3/2, 5/2)^-$, and $3/2^+$ to A^{39} ground state and to levels at 1.27 and 1.52 Mev, respectively (Pe 56).

For jj-coupling computation of the β^- matrix element, Gr 56c.

Also, En 54a.

II. $S^{36}(\alpha,p)Cl^{39}$ $Q_m=-5.7$

Observed, En 54a.

III. $A^{40}(\gamma,p)Cl^{39}$ $Q_m=-12.50$

Threshold, 14.2 ± 0.2 Mev (Ha 50a). This value probably too high because no correction applied for penetration of proton through Coulomb barrier. Q value also measured at $E_\gamma=17.6$ Mev from proportional counter pulse-height analysis as -10.8 ± 0.1 Mev. A second proton group may be present leading to a Cl^{39} level at about 1 Mev (Wi 51).

Cross sections, Mc 54a, Sp 55a.

A^{39}

(Fig. 37)

I. $A^{39}(\beta^-)K^{39}$ $Q_m=0.565$

Half-life, 265 ± 30 yr (Ze 52).

The β^- spectrum end point is determined by lens spectrometer as 565 ± 5 kev. Shape of spectrum is unique first forbidden ($\Delta J=2$, yes) which agrees with the high ft value ($\log ft=9.9$ and $\log(W_0^2-1)\,ft=10.4$) (Br 50). No γ rays observed (Br 50, An 52).

II. $A^{38}(n,\gamma)A^{39}$ $Q_m=6.607$

Cross section, Hu 55c.

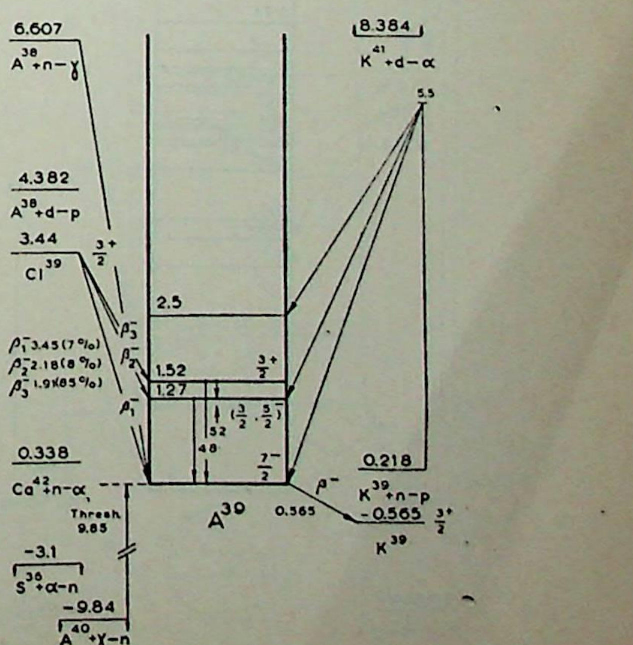


Fig. 37.

III. $Cl^{39}(\beta^-)A^{39}$ See Cl^{39}

IV. $K^{39}(n,p)A^{39}$ $Q_m=0.218$

A KI scintillation crystal bombarded with 2.0–5.5 Mev neutrons; from pulse-height analysis, proton groups observed leading to A^{39} levels at 1.24 ± 0.05 and 2.46 ± 0.1 Mev, if ground-state $Q=0.20$ Mev is assumed (Sc 56a).

V. $A^{40}(\gamma,n)A^{39}$ $Q_m=-9.84$

Threshold, 9.85 ± 0.15 Mev (Ha 54). Cross section, Fe 54, Mc 54a.

VI. The following reactions leading to A^{39} have not been observed: $S^{36}(\alpha,n)A^{39}$ ($Q_m=-3.1$), $A^{38}(d,p)A^{39}$ ($Q_m=4.382$), $K^{41}(d,\alpha)A^{39}$ ($Q_m=8.384$), $Ca^{42}(n,\alpha)A^{39}$ ($Q_m=0.338$).

K^{39}

(Fig. 38)

I. $A^{36}(\alpha,p)K^{39}$ $Q_m=-1.282$

At $E_\alpha=7.4$ Mev, with enriched targets and proton detection with nuclear emulsions, $Q=-1.28\pm0.03$, -3.78 ± 0.06, and -4.15 ± 0.04 Mev, corresponding to transitions to K^{39} ground state and to levels at 2.50 and 2.87 Mev (Sc 56b).

II. (a) $A^{38}(p,\gamma)K^{39}$ $Q_m=6.389$
(b) $A^{38}(d,n)K^{39}$ $Q_m=4.164$

Not observed.

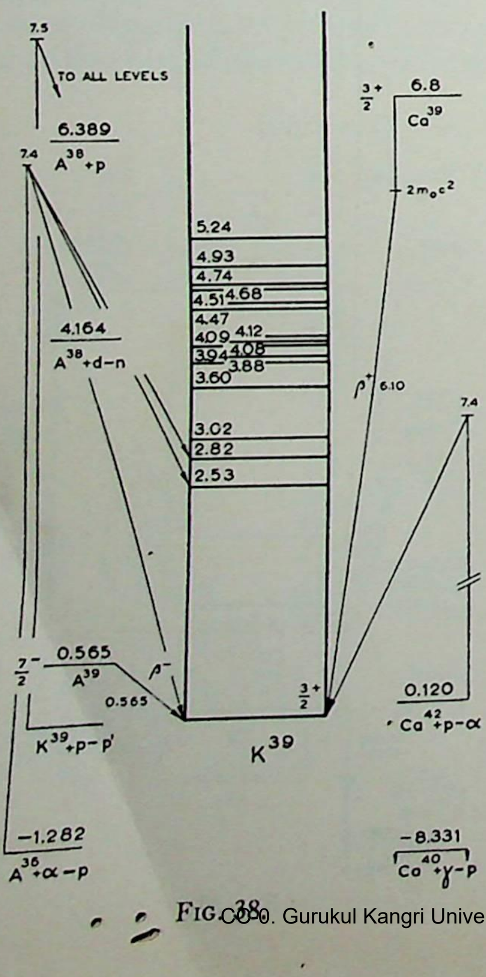


FIG. 38.

III. $A^{39}(\beta^-)K^{39}$ See A^{39}

IV. $K^{39}(p,p')K^{39}$

By magnetic analysis at several angles of observation and at several proton energies up to 7.5 Mev, K^{39} levels observed at 2.526, 2.819, 3.018, 3.600, 3.880, 3.936, 4.080, 4.091, 4.121, 4.472, 4.512, 4.679, 4.735, 4.928, and 5.238 Mev, all ±0.010 Mev (Sp 55, Sp 55b). Absence of levels below $E_x=2$ Mev has been confirmed (Sc 56c).

V. $Ca^{39}(\beta^+)K^{39}$ See Ca^{39}

VI. $Ca^{40}(\gamma,p)K^{39}$ $Q_m=-8.331$

Cross section, Jo 55, Mo 55a.

VII. $Ca^{42}(p,\alpha)K^{39}$ $Q_m=0.120$

By magnetic analysis at $\theta=90°$ and $E_p=6.5$ and 7.4 Mev, the ground-state $Q=0.118\pm0.005$ Mev (Br 56f).

Ca^{39}

(not illustrated)

I. $Ca^{39}(\beta^+)K^{39}$ $Q_m=6.8$

Half-life averaged from four measurements, (En 54a, Kl 54b) 0.93 ± 0.03 sec.

The β^+ end point, determined by scintillation spectrometer, $=6.10\pm0.15$ Mev (Hu 54) and by Al absorption $=6.7\pm0.5$ Mev (Br 53).

Log $ft=3.7$.

II. $A^{36}(\alpha,n)Ca^{39}$ $Q_m=-8.9$

Not observed.

III. $K^{39}(p,n)Ca^{39}$ $Q_m=-7.6$

Not observed.

IV. $Ca^{40}(\gamma,n)Ca^{39}$ $Q_m=-15.9$

Threshold, 15.8 ± 0.1 Mev (Su 53), 15.9 ± 0.4 Mev (Mc 49), and 16.0 ± 0.3 Mev (Be 47).

Cross section, En 54a, Go 54a, Mo 55a.

Cl^{40}

(not illustrated)

I. $Cl^{40}(\beta^-)A^{40}$ $Q_m=7.5$

Half-life, 1.4 min. The β^- spectrum investigated with a scintillation spectrometer; end point, 7.5 Mev. Strong β^- transition with end point between 3.0 and 3.5 Mev also present. Gamma rays of 1.46, 2.75, and 6.0 Mev observed; first two have about equal intensities. Data suggest decay scheme as in Fig. 39. Compatible with assignment of 2^- to Cl^{40} and 0^+, 2^+, and 1^- to ground

state and levels at 1.46 and 6.0 Mev of A^{40}, respectively (Mo 56b).

II. $A^{40}(n,p)Cl^{40}$ $Q_m=-6.7$

Observed, Mo 56b.

A^{40}

(Fig. 39)

I. $A^{40}(e,e')A^{40}$

Magnetic analysis of scattering of 187-Mev electrons shows levels at 1.46 and 2.4 Mev; angular distributions also measured (He 56).

II. $A^{40}(p,p')A^{40}$

Inelastic scattering of protons by A^{40} investigated with 9.5-Mev cyclotron protons and nuclear emulsions for detection of scattered protons. Levels found at 1.48 ± 0.02 Mev (Fr 54), and at 1.48 ± 0.02, 2.22 ± 0.04, (2.66), 3.12 ± 0.03, 3.80 ± 0.03, (4.40), 4.50 ± 0.05, and 4.98 ± 0.05 Mev (Va 56e). Angular distributions of elastically and inelastically scattered protons, (He 47, Fr 54, Hi 55, Bu 56a, Hi 56, Va 56e).

III. $A^{40}(\alpha,\alpha')A^{40}$

Scattering of 19-Mev α particles by argon and angular distributions of groups leading to ground state and to 1.46-Mev level, (Bl 56a).

IV. $K^{40}(EC)A^{40}$ See K^{40}

V. Following reactions leading to A^{40} not observed: $Cl^{37}(\alpha,p)A^{40}$ $(Q_m=-1.61)$, $K^{41}(\gamma,p)A^{40}$ $(Q_m=-7.84)$, and $Ca^{43}(n,\alpha)A^{40}$ $(Q_m=2.25)$.

K^{40}

(Fig. 40)

I. $K^{40}(EC)A^{40}$ $Q_m=1.48$
$K^{40}(\beta^-)Ca^{40}$ $Q_m=1.319$

Most accurate measurement of number of β^- particles emitted per sec per g of potassium yields 27.50 ±0.25 (Mc 56), in excellent agreement with weighted average, 27.4 ± 0.6, of other, less accurate, measurements (En 54a, He 54a, Su 55, Ko 55).

Measurements of β^- end point (En 54a, Ko 55) yield weighted average of 1.328 ± 0.010 Mev; log $ft=18.1$,

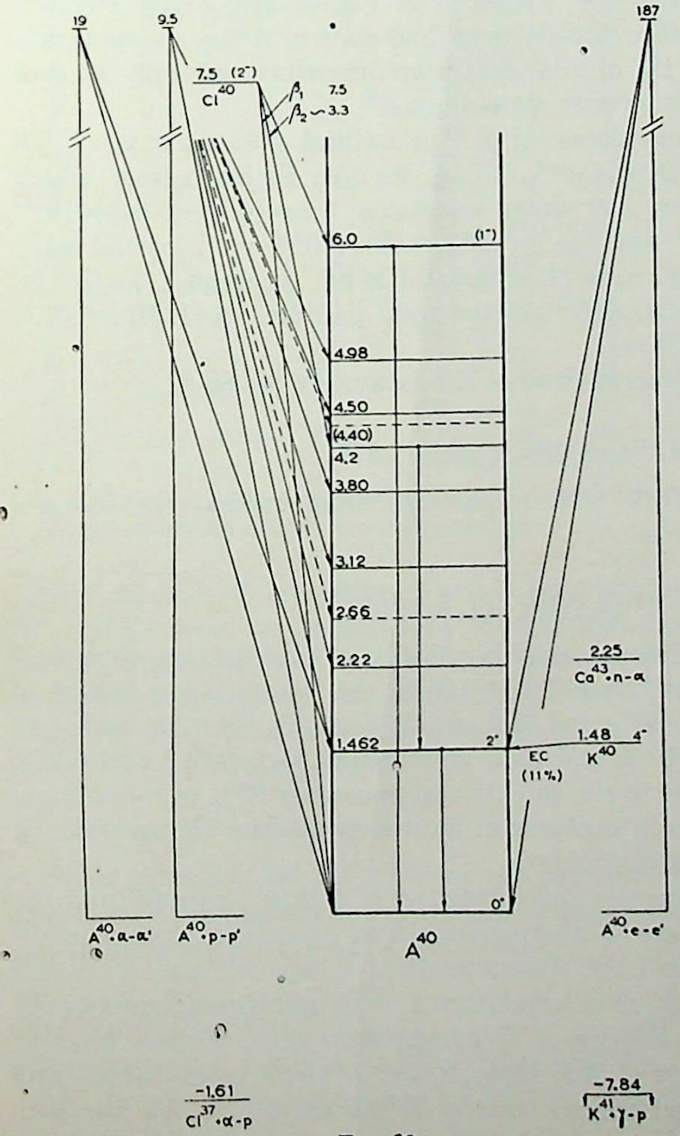


FIG. 39.

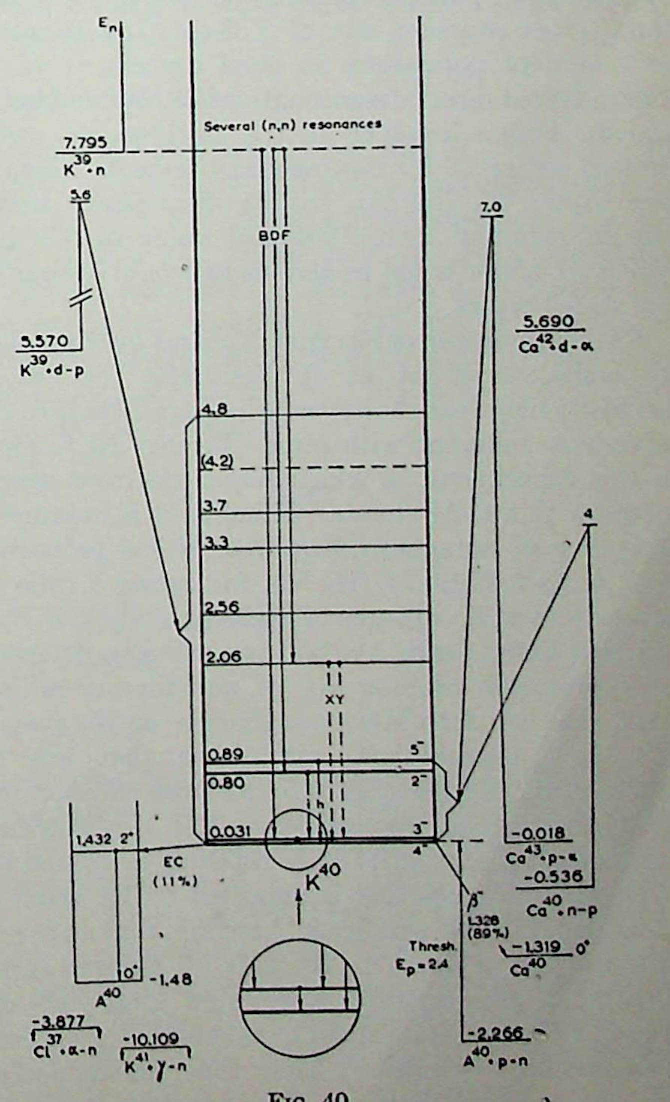


FIG. 40.

and shape of spectrum is unique third forbidden (En 54a, Ko 55), in agreement with assignment of $J=4^-$ to K^{40} (see Ra 53) and $J=0^+$ to Ca^{40}.

Measurement with lowest stated error of number of γ quanta emitted per sec per g of potassium yields 2.96 ± 0.04 (Su 55), while weighted average of six less accurate results (Gl 47, Ah 48, Fa 50, Ho 50, Bu 53, Ba 55b) is 3.42 ± 0.06. This discrepancy discussed below.

Measurements of γ energy (En 54a, Me 56) yield a weighted average of 1.462 ± 0.005 Mev. The $K^{40}-Ca^{40}$ and $K^{40}-A^{40}$ mass differences determined mass-spectroscopically as 1.30 ± 0.07 and 1.49 ± 0.07 Mev, respectively (Jo 52a), indicating that γ ray must follow electron capture. This confirmed by independent experiments (see under A^{40} and Ca^{40}) which show that A^{40} has excited state at 1.48 ± 0.02 Mev, and that first excited state of Ca^{40} is at 3.35 Mev. Coincidences between γ quanta and Auger electrons also confirm assignment of γ ray to A^{40}(Pa 52, Pa 52a).

Most accurate direct measurements of ratio of number of γ quanta to number of electrons yield 0.124 ±0.002 and 0.121 ± 0.004 (Mc 56); less accurate measurements (En 54a) were lower. Adopting 0.123 ± 0.002, and combining this with adopted value of 27.5 ± 0.2 for number of β^- particles emitted per sec per g of potassium, gives emission rate of 3.38 ± 0.07 γ quanta per sec per g of potassium, in good agreement with all except one of direct determinations of this number (see above). Errors involved in the thick-source method, devised especially for routine measurements, amount to five to ten % (Su 55), so the discrepancy between quoted result of 2.96 ± 0.04 and value of 3.38 ± 0.05 which we adopt is not as serious as would appear from the stated errors.

Ratio of number of electron captures over number of β^- emissions given as 0.135 ± 0.040 (Sa 50a, also En 54a); not clear, however, whether L capture could have been detected with same efficiency as K capture in this experiment, in which Auger electrons were detected with a G-M counter. Number of K captures per sec per g of potassium measured with a proportional counter as 1.42 ± 0.23 (He 54), indicating a ratio of L captures over K captures of approximately 1.4.

Upper limits for rate of positron emission determined by comparison of intensity of annihilation radiation with that of 1.46-Mev γ radiation or β^- radiation (En 54a). Appears well established that less than 0.05% of the K^{40} decay goes by positron emission to A^{40} ground state. Since approximately 1.48 Mev is available for positron decay, ratio of K captures to ground state over positron emissions is expected to be about 200 (Mo 51a). It is not excluded, therefore, that an appreciable fraction of decay to A^{40} goes by electron capture directly to ground state. Results on total yield of K Auger electrons (see above), as well as geochemical measurements (see below), show that this fraction cannot be very large, however. This conclusion supported by theoretical arguments: the ft for β^+ and β^- transitions to ground states of A^{40} and Ca^{40} expected to be same, hence, intensity ratio of β^+ and β^- transitions expected to be of order of $1:10^4$ (Mo 51a).

Measurements of amount of argon in minerals of known age which contain potassium, and mass spectrometer studies of relative amounts of A^{40}, K^{40}, and Ca^{40} in such minerals (En 54a, Sh 54a, Wa 55c, We 56) yield values for electron capture to β^- emission branching ratio which generally come out lower than value deduced from counting experiments (however, In 50). Recent review of geochemical experiments (We 56), however, shows several older determinations need upward revision and that geological evidence gives a branching ratio of 0.117 ± 0.015, in good agreement with value, 0.123 ± 0.002, deduced from counting experiments on assuming every electron capture is followed by a 1.46-Mev γ transition.

Foregoing discussions may be summarized as follows: it is definitely established that K^{40} decays by 1.33-Mev β^- transition to ground state of Ca^{40} and by electron capture to first excited state of A^{40}, which in turn decays to ground state by a 1.46-Mev γ transition. Ratio of β^- emissions to γ emissions is 89.0 to 11.0. It is not excluded that a few % of the decays go by electron capture directly to ground state of A^{40}; γ-emission rate of 11% of total decay rate, therefore, sets only a lower limit for rate of decay to A^{40}.

Abundance of K^{40} in natural potassium is $(1.178\pm0.004)\times10^{-4}$ (Ni 50, Re 52a, Re 56c), from which follow the decay constants $\lambda_{\beta^-}=(4.78\pm0.04)\times10^{-10}$ yr^{-1} and $\lambda_{EC}\geq(5.88\pm0.09)\times10^{-11}$ yr^{-1}, partial half-lives $\tau_{\beta^-}=(1.449\pm0.013)\times10^{9}$ yr and $\tau_{EC}\leq(1.18\pm0.02)\times10^{10}$ yr, and total half-life $\tau\leq(1.291\pm0.013)\times10^{9}$ yr.

Other reviews of K^{40} decay, We 56, Sh 57.

II. $Cl^{37}(\alpha,n)K^{40}$ $Q_m=-3.877$

Work with α particles from radioactive sources, En 54a.

III. $K^{39}(n,\gamma)K^{40}$ $Q_m=7.795$

Thermal neutron absorption cross section of natural potassium is 1.97 ± 0.06 b; the isotopic cross section of K^{39}, K^{40}, and K^{41} are 1.87 ± 0.15, 70 ± 20, and 1.19 ±0.10 b, and their abundances 93.1, 0.012, and 6.9%, respectively, so that approximately 95% of the thermal neutron captures in natural potassium should occur in K^{39} (Hu 55c).

Energies and intensities of γ rays resulting from the capture of thermal neutrons in natural potassium in Table LIII. Assignments of γ rays A'' and A' to K^{41} based upon experiments with potassium enriched in K^{40}. Energy of C corresponds with neutron binding energy of K^{42}. Lines B and D go to ground state and second excited state of K^{40}, while h and i fit between low-lying levels of K^{40}; several others must be assigned

TABLE LIII. Gamma rays from thermal neutron capture in potassium.

Reference:	Ki 52, Ba 53		Ad 56		
Method:	Pair spectrometer		Compton spectrometer		
Line	$E\gamma$ (Mev)	Intensity [a]	$E\gamma$ (Mev)	Intensity [a]	Final nucleus
A''	9.39 ± 0.06	(0.02)			K^{41}
A'	8.45 ± 0.02	(0.1)			K^{41}
B	7.757 ± 0.008	(3.5)	7.763 ± 0.010	(4.4)	K^{40}
C	7.34 ± 0.02	(0.1)	7.32 ± 0.03	(0.2)	K^{42}
D	6.994 ± 0.007	(1.3)	7.000 ± 0.015	(1.6)	K^{40}
E	6.31 ± 0.06	(0.3)			
F	5.740 ± 0.012	(6)	5.725 ± 0.015	(11)	K^{40}
F'	5.66 ± 0.02	(4)	(5.65 ± 0.03)	(3)	
F''	5.50 ± 0.02	(2.5)			
G	5.38 ± 0.03	(6)	5.40 ± 0.02	(6.4)	
G'	5.18 ± 0.02	(2)	5.25 ± 0.05	(2.5)	
H	5.06 ± 0.02	(3)	5.02 ± 0.03	(4.5)	
H'			4.81 ± 0.04	(1.5)	
H''			4.70 ± 0.03	(1)	
H'''			4.50 ± 0.04	(1)	
I	4.39 ± 0.03	(4)	4.39 ± 0.02	(5)	
J	4.18 ± 0.05	(6)	4.11 ± 0.03	(4)	
K	3.92 ± 0.05	(5)	3.97 ± 0.03	(5)	
K'			3.81 ± 0.05	(2)	
L	3.67 ± 0.05	(8)	3.70 ± 0.03	(4)	
	Br 56e Two-crystal scint. spectrometer				
M			3.60 ± 0.03	(4)	
N			3.55 ± 0.05	(2)	
O	3.45 ± 0.05	(8)	3.40 ± 0.05	(4)	
P			(3.15 ± 0.05)	(1.5)	
Q			3.05 ± 0.03	(3)	
R	2.80 ± 0.03	(6)	2.75 ± 0.04	(4)	
S			(2.60 ± 0.04)	(4)	
T			(2.55 ± 0.03)	(4)	
U			(2.42 ± 0.03)	(4)	
V			(2.37 ± 0.03)	(2)	
W			(2.30 ± 0.03)	(3)	
X			2.06 ± 0.01	(9)	K^{40}
Y	2.03 ± 0.03	(13)	2.02 ± 0.02	(7)	K^{40}
Z			1.95 ± 0.02	(4)	
a			1.85 ± 0.02	(3)	
b			1.75 ± 0.02	(3)	
c	1.61 ± 0.03	(5)	1.61 ± 0.01	(13)	K^{40}
d			1.51 ± 0.01	(5)	
e			1.40 ± 0.02	(2)	
f			1.27 ± 0.02	(2)	
g	1.19 ± 0.03	(8)	1.18 ± 0.02	(7)	K^{40}
h			0.90 ± 0.02	(2)	K^{40}
i	0.77 ± 0.03	(24)	0.77 ± 0.01	(26)	K^{40}
j			0.625 ± 0.010	(3)	

[a] In photons per 100 captures in natural potassium.

to K^{40} because of intensities. Further experiments on $K^{39}(d,p)K^{40}$ reaction are necessary for determination of energy levels of K^{40}, without which it is impossible to to make unambiguous assignments of other γ rays. Proposed level scheme which accounts for several of observed γ transitions, Ad 56.

IV. $K^{39}(n,n)K^{39}$ $E_b=7.795$

Cross sections, Hu 55c, Jo 55a, To 55; nine resonances found at neutron energies below 110 kev (To 55), several more above 110 kev (Hu 55c).

V. $K^{39}(d,p)K^{40}$ $Q_m=5.570$

Ground-state Q value $=5.506 \pm 0.010$ Mev by magnetic analysis at E_d from 4.8 to 5.7 Mev; excited states found at 0.032 ± 0.002, 0.800 ± 0.010, and 0.893 ± 0.010 Mev (Bu 53a). With range measurements at $E_d=3.9$ Mev, partly resolved levels found at 2.01 ± 0.03, 2.56 ± 0.05, and 3.3, 3.7, (4.2), and 4.8, all ± 0.1 Mev (Sa 50).

VI. $A^{40}(p,n)K^{40}$ $Q_m=-2.266$

Threshold of $E_p=2.4$ Mev, (Ri 48).

VII. $Ca^{40}(n,p\gamma)K^{40}$ $Q_m=-0.536$

Gamma rays of 0.030 ± 0.002, 0.767 ± 0.007, and 0.877 ± 0.017 Mev found from bombardment of calcium with 4-Mev neutrons. They represent transitions in K^{40} from (1) to (0), from (2) to (1), and from (3) to (0) or (1), respectively (Da 56c).

VIII. $K^{41}(\gamma,n)K^{40}$ $Q_m=-10.109$

Not observed.

IX. $Ca^{42}(d,\alpha)K^{40}$ $Q_m=5.690$

Not observed.

X. $Ca^{43}(p,\alpha)K^{40}$ $Q_m=-0.018$

Ground-state Q value $=-0.014 \pm 0.008$ Mev by magnetic analysis at $E_p=6.5$ and 7.0 Mev. Groups of α particles leading to first three excited states of K^{40} observed (Br 56f).

GENERAL REMARKS

Ground state of K^{40} has $J=4$ (Ra 53). Together with first three excited states, it forms a multiplet with spins 4^-, 3^-, 2^-, and 5^-, resulting from coupling of a $f_{7/2}$ neutron with a $d_{3/2}$ proton hole (En 54a). See under Cl^{38}, general remarks, for discussion of a relation between excitation energies in Cl^{38} and K^{40} based on jj coupling. Theoretical discussions of K^{40} ground state configuration, Ku 53, De 53a, Hi 54, Ma 54, Sc 54c, Ta 54a, Pa 56a, Pa 57a.

Ca^{40}

(Fig. 41)

I. $K^{39}(d,n)Ca^{40}$ $Q_m=6.106$

Not observed.

II. $K^{39}(p,\gamma)Ca^{40}$ $Q_m=8.331$

Yield curve measured in region 0.5–1.15 Mev, showing resonances at $E_p=0.883$, 0.925, 0.980, and 1.150 Mev, all ± 10 kev (De 57), and in region 0.8–1.8 Mev, showing resonances at 1.120, 1.300, 1.338, and 1.566 Mev, all ± 5 kev (To 57). First four resonances decay mainly to ground state; at 1.150-Mev resonance, a direct ground-state transition and cascades through the 3.35- and 3.73-Mev levels are found. Angular distribution of 9.29-Mev γ ray indicates this is an E1 transition. High-energy γ rays at 1.150-Mev resonance measured

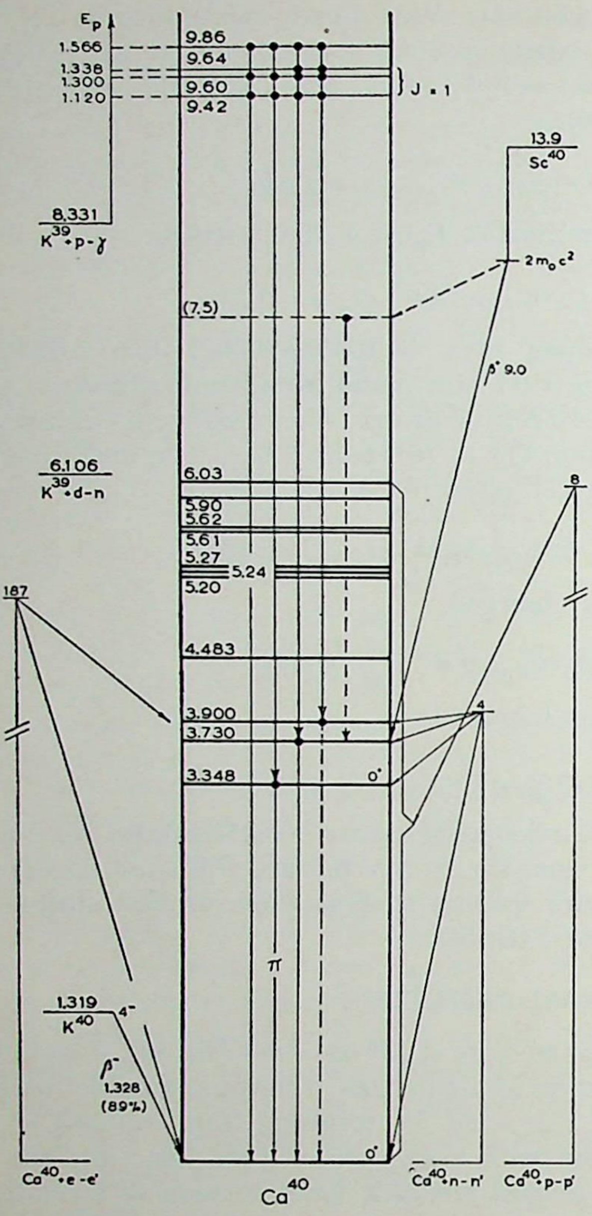


FIG. 41.

with a three-crystal pair spectrometer, indicating $Q=8.17\pm0.03$ Mev, (De 57).

The 1120-, 1300-, and 1566-kev resonances decay with strong ground-state transitions, weak transitions to lowest three levels [those to (2) and (3) unresolved], and perhaps transitions to level (4). The 1338-kev resonance decays to level (2) and (3). Angular distribution measurements yield $J=1$ to the 1120- and 1300-kev resonances, (To 57).

III. $K^{40}(\beta^-)Ca^{40}$ See K^{40}

IV. $Ca^{40}(e,e')Ca^{40}$

Magnetic analysis of scattering of 183-Mev electrons by calcium shows unresolved group leading to levels at 3.73 and 3.90 Mev, also elastically scattered electrons. Angular distributions giving charge distribution, (Ha 56b). Inelastic group leads to 3.73-Mev level, which probably has $J=3^-$ (He 56).

TABLE LIV. Excited states of Ca^{40} (energy in Mev) from $Ca^{40}(p,p')Ca^{40}$ (Br 56b).

3.348±0.004	5.202±0.008	5.621±0.008
3.730±0.004	5.241±0.006	5.901±0.008
3.900±0.004	5.272±0.006	6.029±0.008
4.483±0.005	5.606±0.008	

V. $Ca^{40}(n,n')Ca^{40}$

Gamma rays of 3.74 ± 0.04 and 3.9 ± 0.1 Mev, in addition to annihilation radiation, result from inelastic scattering of 4-Mev neutrons by calcium. Threshold for production of positrons is $E_n=3.36\pm0.05$ Mev (c.m.), showing that first excited state decays by pair emission (Da 56c). Half-life of this state measured with a pulsed neutron source; preliminary value of 1.4×10^{-9} sec obtained, (Kl 57). Elastic neutron scattering, We 56a, Cu 56.

VI. $Ca^{40}(p,p')Ca^{40}$

Magnetic analysis of inelastic scattering of 6- to 8-Mev protons by Ca^{40} shows excited states in Table LIV (Br 56b). With a magnetic spectrometer, 3.35-Mev level found to decay by pair emission; no γ ray found. Hence, its spin and parity must be 0^+ (Be 55b).

Angular distribution of an unresolved proton group to levels at 3.73 and 3.90 Mev, taken at $E_p=7.7$ Mev, nearly isotropic (Ha 52c).

VII. $Sc^{40}(\beta^-)Ca^{40}$ $Q_m=13.9$

Half-life is 0.22 ± 0.3 sec. Maximum positron energy, measured with a scintillation spectrometer, is 9.0 ± 0.4 Mev, which, assuming a branching of 100%, gives log ft=4.1. A γ ray of 3.75 ± 0.04 Mev, as measured with a scintillation spectrometer, accompanies the decay (Gl 55). Possibility that a 5-Mev super-allowed β^+ branch leads to lowest $T=1$ state, expected at 7.5 Mev, not excluded by experimental data (Bo 55). No delayed α particles observed (Gl 55). Theoretical discussion of position of $T=1$ states in self-conjugated nuclei, Wi 56a.

A^{41}

(Fig. 42)

I. $A^{41}(\beta^-)K^{41}$ $Q_m=2.524$

Half-life averaged from four measurements (En 54a, Ha 51d) is 109.6 ± 0.4 min. Also An 54.

The β^- decay proceeds by two branches. End point of ground-state transition measured by magnetic spectrometer as 2.48 ± 0.04 Mev (intensity 0.88%) which yields log $ft=8.5$ and $\log(W_0^2-1)\ ft=10.0$. Shape of spectrum is unique once-forbidden ($\Delta J=2$, yes) (Sc 56). End point of main branch measured by magnetic spectrometer as 1.199 ± 0.008 Mev (Sc 56), and 1.245 ± 0.005 Mev (Br 50c). Spectrum has allowed shape. Log $ft=5.1$.

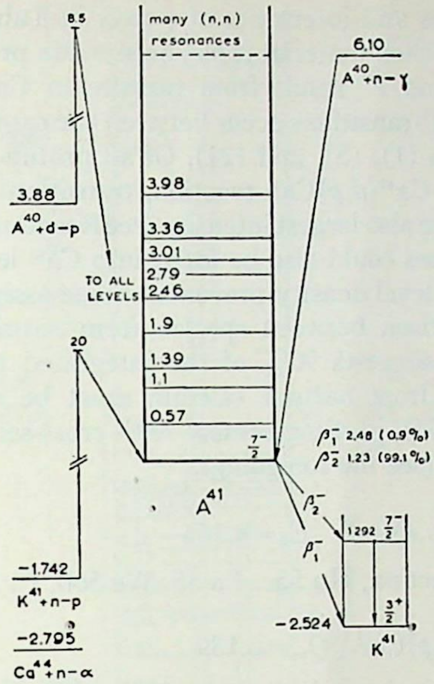


FIG. 42.

Energy of γ ray following main β^- branch measured by scintillation spectrometer as 1.298 ± 0.010 Mev (Kl 55), and 1.290 ± 0.005 Mev (Sc 56).

By β-γ delayed coincidence measurements half-life of 1.3-Mev level in K^{41} is established as $(6.7\pm0.5)\times10^{-9}$ sec (El 52a) and 6.6×10^{-9} sec (En 53) which agrees with theoretical value for M2 transitions.

Also En 54a.

For jj-coupling computation of β^- matrix element, Gr 56c.

II. (a) $A^{40}(n,\gamma)A^{41}$ $E_b=6.10$
(b) $A^{40}(n,n)A^{40}$ $E_b=6.10$

Cross section, Hu 55c.

III. $A^{40}(d,p)A^{41}$ $Q_m=3.88$

Results of Q-value and angular distribution measurements in Table LV. At $E_d=3.9$ Mev, by Al absorption, ground-state Q value $=3.80\pm0.06$ Mev (Sa 49). Theoretical analysis of A^{41} levels, Ra 56a.

Also Fu 54, Fr 55, En 54a.

TABLE LV. The $A^{40}(d,p)A^{41}$ reaction.

Reference:	Da 49	Gi 52		Bu 56d	
E_d (Mev):	3.4	7.8		8.5	
Method:	Al abs.	Nucl. emuls.		Magn. anal.	
Q_0 (Mev):	3.84 ± 0.03	3.90 ± 0.08		...[b]	
	E_x (Mev)	E_x (Mev)	l_n	E_x (Mev)	l_n
		0	3	0	3
	0.66 ± 0.05 [a]	0.50 ± 0.08 [a]		0.57 ± 0.05	1
	1.21 ± 0.04 [a]			1.11 ± 0.1	
	1.34 ± 0.04 [a]	1.34 ± 0.08 [a]	0, (1)	1.39 ± 0.05	1
	1.94 ± 0.04 [a]			1.93 ± 0.1	
	2.27 ± 0.05 [a]			2.46 ± 0.05	1
	2.80 ± 0.05 [a]			2.79 ± 0.05	1
	3.29 ± 0.03 [a]			3.01 ± 0.05	1
	3.69 ± 0.05 [a]			3.36 ± 0.05	1
	4.01 ± 0.05 [a]			3.98 ± 0.05	1

[a] In original paper errors were only assigned to Q values. Same errors have been used here for the corresponding excitation energies.
[b] A ground-state Q value of 3.90 Mev is assumed.

IV. $K^{41}(n,p)A^{41}$ $Q_m=-1.742$

Cross section, En 54a.

V. $Ca^{44}(n,\alpha)A^{41}$ $Q_m=-2.795$

Not observed.

K^{41}

(Fig. 43)

I. $A^{40}(p,\gamma)K^{41}$ $Q_m=7.84$

Thin targets enriched in A^{40} content bombarded with protons from 0.5 to 1.8 Mev. Resonances in γ-ray yield found (Br 48a) at

E_p	900	1050	1080	1100	1235 kev
Rel. int.	0.2	0.4	0.5	1.0	0.5

II. $A^{40}(d,n)K^{41}$ $Q_m=5.62$

Neutron detection with a double proportional counter, at $E_d=3.2$ Mev, gives ground-state Q value $=5.97\pm0.25$ Mev and levels at 1.34 ± 0.15, 3.10, and 4.0 Mev (Wo 50).

III. $K^{40}(n,\gamma)K^{41}$ $Q_m=10.109$

Cross section, Hu 55c.

Two γ rays A'' and A' of Table LIII produced by neutron capture in potassium assigned to K^{41} on account of their energies and from experiments with targets enriched in K^{40}. If originating from the capturing state, they would proceed to K^{41} levels at 0.72 and 1.66 Mev (Ba 53, Ki 52).

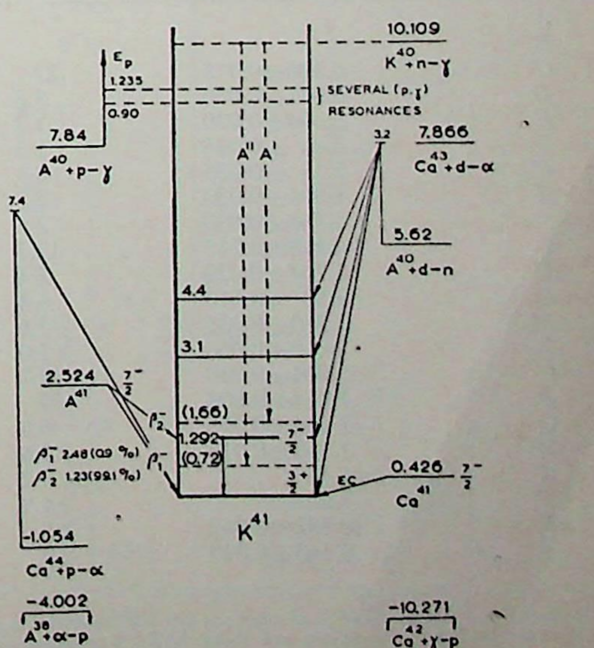


FIG. 43.

IV. $A^{41}(\beta^-)K^{41}$ See A^{41}

V. $Ca^{41}(EC)K^{41}$ See Ca^{41}

VI. $Ca^{44}(p,\alpha)K^{41}$ $Q_m = -1.054$

By magnetic analysis at $\theta = 90°$, $E_p = 6.5$ and 7.4 Mev, ground-state Q value $= -1.057 \pm 0.010$ Mev (Br 56f).

VII. Following reactions leading to K^{41} have not been observed: $A^{38}(\alpha,p)K^{41}$ ($Q_m = -4.002$), $Ca^{42}(\gamma,p)K^{41}$ ($Q_m = -10.271$), $Ca^{43}(d,\alpha)K^{41}$ ($Q_m = 7.866$).

Ca^{41}

(Fig. 44)

I. $Ca^{41}(EC)K^{41}$ $Q_m = 0.426$

Half-life is $(1.1 \pm 0.3) \times 10^5$ yr, assuming a Ca^{40} thermal neutron capture cross section of 0.22 b (Br 51a, Br 53a).

II. $A^{38}(\alpha,n)Ca^{41}$ $Q_m = -5.211$

Not observed.

III. $Ca^{40}(n,\gamma)Ca^{41}$ $Q_m = 8.364$

Thermal-neutron absorption cross section of natural calcium is 0.43 ± 0.02 b; main contributions from Ca^{40} and Ca^{42}, with cross sections of 0.22 ± 0.04 and 40 ± 3 b, and abundances of 97% and 0.64%, respectively, indicating that these isotopes should contribute about equally to cross section of natural calcium (Hu 55c).

Energies and intensities of γ rays in Table LVI. By using enriched material it has been made probable that lines F and F' result from capture in Ca^{42} (Ad 56). Strong E1 transitions occur between the capturing state and levels (1), (3), and (21). Of all proton groups observed in $Ca^{40}(d,p)Ca^{41}$ reaction, transitions to these p states have also largest intensity (see Reaction V). Many weaker lines could also be fitted into Ca^{41} level scheme but great level density prevents unique assignment.

Comparison between spectra from natural calcium and Ca^{40} suggests 90% of the integrated intensity of spectrum from natural calcium must be ascribed to Ca^{40} (Ad 56), in disagreement with cross-section measurements (see the foregoing).

IV. $Ca^{40}(n,n)Ca^{40}$ $E_b = 8.364$

Cross section, Hu 55c, Pa 55, We 56b, Cr 55.

V. $Ca^{40}(d,p)Ca^{41}$ $Q_m = 6.139$

By magnetic analysis at $\theta = 90°$ and at several bombarding energies between 2.5 and 7.0 Mev, ground-state Q value $= 6.140 \pm 0.009$ Mev (Br 56f, Br 56g). Levels in Ca^{41} in Table LVII. Angular distributions also measured with a high-resolution magnetic analyzer at $E_d = 7.0$ Mev (Bo 56, Bo 56a); l_n assignments in Table LVII. Transitions to Ca^{41} ground state and to levels (1), (3), and (21) very intense. Level density above $E_x = 4.2$ Mev increases very fast. Two groups leading to levels at 4.76 and 5.72 Mev, both with $l_n = 2$, observed in older low-resolution work (Ho 53c).

Also En 54a.

Discussion of reduced widths, Fu 54, Ra 56.

TABLE LVI. Gamma rays from thermal neutron capture in natural calcium.

Reference:	Ad 56		Ki 52		
Method:	Magn. Compton spectrometer		Magn. pair spectrometer		Probable transition
Line	Energy (Mev)	Intensity [a]	Energy (Mev)	Intensity [a]	in Ca^{41} [b]
A			7.83 ± 0.05	1	
B			7.43 ± 0.05	1.4	
C	6.406 ± 0.015	22	6.42 ± 0.03	83 (80)[c]	$(c) \to (1)$
D	5.904 ± 0.030	3.8	5.89 ± 0.03	11 (12)[c]	$(c) \to (3)$
E	5.696 ± 0.030	1.2	5.66 ± 0.06	3	
F	5.50 ± 0.035	1.2	5.49 ± 0.05	4	$Ca^{42}(n,\gamma)Ca^{43}$
F'	5.15 ± 0.035	0.9			$Ca^{42}(n,\gamma)Ca^{43}$
H	4.944 ± 0.030	2.3	4.95 ± 0.03	8	
I	4.764 ± 0.030	2.5	4.76 ± 0.03	6	
J	4.418 ± 0.015	12.3	4.45 ± 0.05	30	$(c) \to (21)$
J'	3.762 ± 0.020	1.8			
L	3.60 ± 0.010	6.4	3.62 ± 0.05	16	
M	2.810 ± 0.035	$\geqslant 3.6$	Br 56e		
N	2.660 ± 0.050	$\geqslant 2.0$	Two-crystal scint. spectrometer		$(21) \to (1), (2) \to (0)$
O	2.004 ± 0.010	12.7			
P	1.944 ± 0.008	39	1.93 ± 0.03	45	$(1) \to (0)$
Q	1.844 ± 0.015	6.4			
R	1.790 ± 0.015	$\geqslant 3.6$			
S	~ 1.48	$\geqslant 3.2$			$(21) \to (3)$
T	~ 1.2	$\geqslant 4.8$			
U	0.532 ± 0.010	$\geqslant 5$			$(3) \to (1)$
V	0.463 ± 0.010	$\geqslant 9$	0.48 ± 0.05	15	$(3) \to (2)$

[a] Intensities given as number of photons per 100 captures in natural calcium.
[b] For numbering of Ca^{41} levels, see Table LVII. Capturing state is denoted by (c).
[c] Intensities in brackets are numbers of photons per 100 captures in Ca^{40}, recalculated with assumption that 50% of the captures in natural calcium occur in Ca^{40} (Ki 54a).

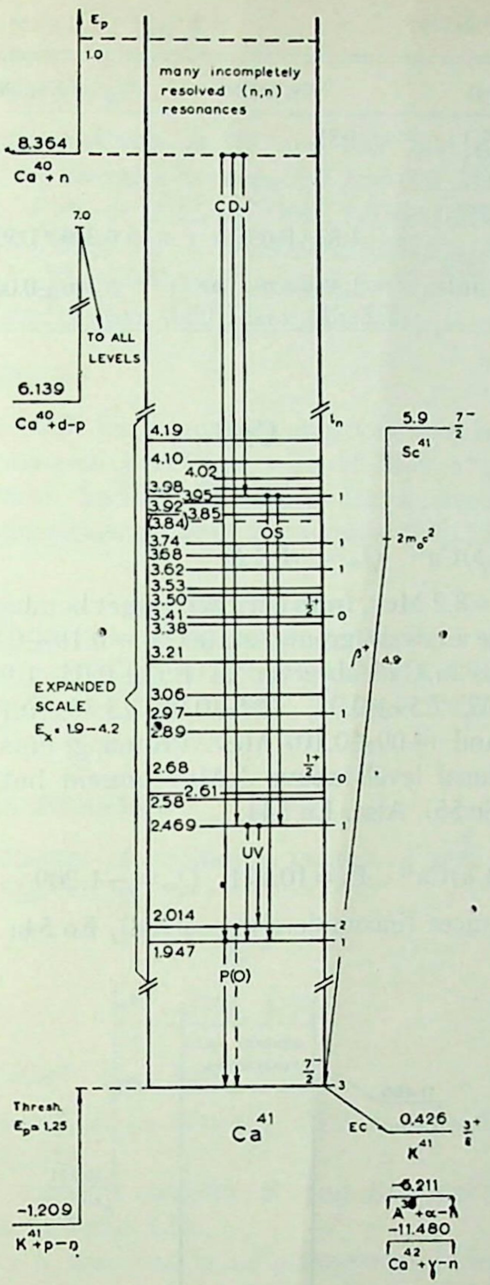


FIG. 44.

VI. $K^{41}(p,n)Ca^{41}$ $Q_m=-1.209$

Threshold for neutron emission, $E_p=1.25\pm0.02$ Mev (Ri 50). Resonances, En 54a.

TABLE LVII. Levels in Ca^{41} from $Ca^{40}(d,p)Ca^{41}$ reaction.

Reference:	Br 56f, Br 56g E_x (Mev)	Bo 56, Bo 56a l_n		Br 56f, Br 56g E_x (Mev)	Bo 56, Bo 56a l_n
(0)	0	3	(13)	3.500±0.007	
(1)	1.947±0.004	1	(14)	3.531±0.007	
(2)	2.014±0.005	2	(15)	3.619±0.007	1
(3)	2.469±0.005	1	(16)	3.682±0.007	
(4)	2.584±0.006		(17)	3.736±0.007	1
(5)	2.612±0.006		(18)	(3.836±0.007)	
(6)	2.677±0.006	0	(19)	3.854±0.007	
(7)	2.890±0.006		(20)	3.921±0.007	
(8)	2.967±0.006	1	(21)	3.950±0.007	1
(9)	3.056±0.006		(22)	3.982±0.007	
(10)	3.206±0.006		(23)	4.023±0.007	
(11)	3.375±0.007		(24)	4.101±0.007	
(12)	3.405±0.007	0	(25)	4.194±0.007	

VII. $Sc^{41}(\beta^+)Ca^{41}$ $Q_m=5.9$

Half-life, 0.87 ± 0.03 sec (El 41), and 0.873 sec (Ma 52a). The β^+ end point by cloud chamber is 4.94 ± 0.07 Mev (El 41). Log $ft=3.4$, indicating superallowed transition.

VIII. $Ca^{42}(\gamma,n)Ca^{41}$ $Q_m=-11.480$

Not observed.

A^{42}

(not illustrated)

This nucleus has been produced by successive capture of two neutrons in A^{40} in a nuclear reactor. Presence detected by observation of 12.5-hr K^{42} daughter activity in milkings from irradiated argon gas. From the fact that the milkings over a period of 400 days did not show a decrease in activity of more than 20%, half-life of $A^{42}>3.5$ yr and thermal neutron activation cross section of $A^{41}>0.06$ b (Ka 52a).

K^{42}

(Fig. 45)

I. $K^{42}(\beta^-)Ca^{42}$ $Q_m=3.528$

Best measurement of half-life, 12.516 ± 0.007 hr (Bu 53), in good agreement with other less accurate measurements (En 54a).

The β^- spectrum is complex; β^- and γ energies and intensities in Table LVIII. Older measurements, En 54a.

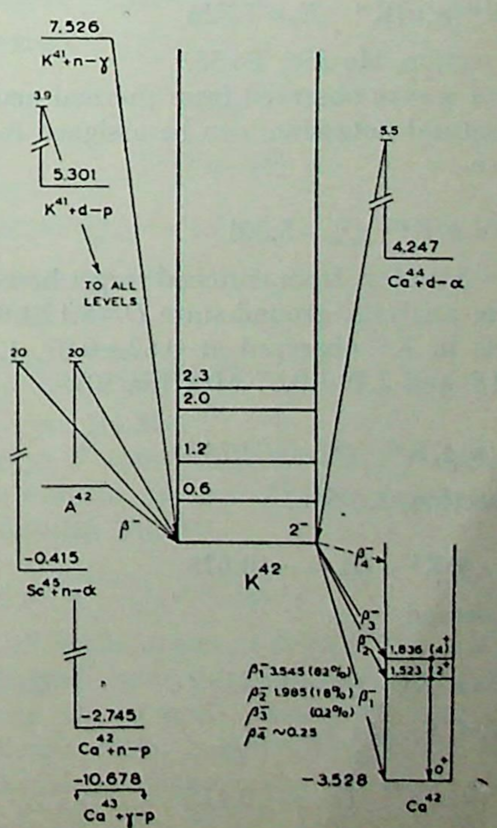


FIG. 45.

TABLE LVIII. The $K^{42}(\beta^-)Ca^{42}$ decay.

Reference	Method	E_{β_1} (Mev)	E_{β_2} (Mev)	E_{γ_1} (Mev)	E_{γ_2} (Mev)
Si 47	magn. spectrom.	3.58 ±0.07 (75%)	2.04 (25%)	1.51	
Si 47a	β-γ coinc. spectrom.		1.92 (16%)		
Ko 54	magn. spectrom.	3.56 (81.8%)			
	β-γ coinc. spectrom.		1.97 (18.2%)		
La 54	scint. spectrom.			1.52 (100%)	0.309 (1.5%)
Em 55	abs. β and γ counting		(10.8±0.6)%		
Po 56	magn. spectrom.	3.545±0.010	1.985±0.010 [a]	1.53±0.01(100%)	0.320±0.005 (0.8%)

[a] A third β^- branch with ~0.25-Mev end point may be present.

Shape of high-energy β^- spectrum and ft value ($\log ft = 7.9$ and $\log(W_0^2-1)\, ft = 9.7$) in agreement with a $\Delta J = 2$, yes, assignment which characterizes K^{42} as $J = 2^-$ (Si 47, Sh 49, Ko 54, Po 56). Direct measurement by magnetic resonance in atomic beam yields also $J = 2$ (Be 53). Assignment of $J = 2^+$ to 1.523-Mev level in Ca^{42} in agreement with ft value of low-energy β^- branch (Sh 49) ($\log ft = 7.4$) and with measured β-γ angular correlation (Be 50, St 51b).

No polarization found of γ rays emitted at 90° to β rays, although some polarization is predicted from measured β-γ angular correlation (Ha 53a).

Measurement of angular correlation of the 0.31- and 1.52-Mev γ rays suggests assignments of $J = 0$, 2, and 4 to Ca^{42} ground state and levels at 1.523 and 1.836 Mev, respectively (Ca 54). However, intensity ratio observed disagrees with that from more direct measurements (Table LVII).

Theoretical discussion of the K^{42} spin, De 53a, Sc 54c.

II. (a) $K^{41}(n,\gamma)K^{42}$ $E_b = 7.526$
(b) $K^{41}(n,n)K^{41}$ $E_b = 7.526$

Cross section, Hu 55c, To 55.

None of γ rays observed from thermal neutron capture in natural potassium can be assigned to K^{41} with confidence.

III. $K^{41}(d,p)K^{42}$ $Q_m = 5.301$

At $E_d = 3.9$ Mev, from enriched target bombardment and range analysis, ground-state $Q = 5.12 \pm 0.10$ Mev, and levels in K^{42} observed at 0.62 ± 0.07, 1.18 ± 0.07, 1.97 ± 0.15, and 2.29 ± 0.07 Mev (Sa 50).

IV. $Ca^{42}(n,p)K^{42}$ $Q_m = -2.745$

Cross section, En 54a.

V. $Ca^{43}(\gamma,p)K^{42}$ $Q_m = -10.678$

Not observed.

VI. $Ca^{44}(d,\alpha)K^{42}$ $Q_m = 4.247$

Observed, En 54a.

VII. $Sc^{45}(n,\alpha)K^{42}$ $Q_m = -0.415$

Observed, En 54a.

Ca^{42}

(Fig. 46)

I. $K^{39}(\alpha,p)Ca^{42}$ $Q_m = -0.120$

At $E_\alpha = 8.2$ Mev, from enriched target bombardments and range analysis, ground-state $Q = -0.19 \pm 0.07$ Mev, and levels in Ca^{42} observed at 1.51 ± 0.05, 1.95 ± 0.07, 2.29 ± 0.05, 2.59 ± 0.07, 3.02 ± 0.05, (3.30 ± 0.10), 3.75 ± 0.07, and (4.09 ± 0.10) Mev. Proton groups leading to additional levels above 3 Mev present but not resolved (Sc 55). Also, En 54a.

II. $K^{41}(p,n)Ca^{41}$ $E_b = 10.271$ $Q_m = -1.209$

Resonances (incompletely resolved), En 54a.

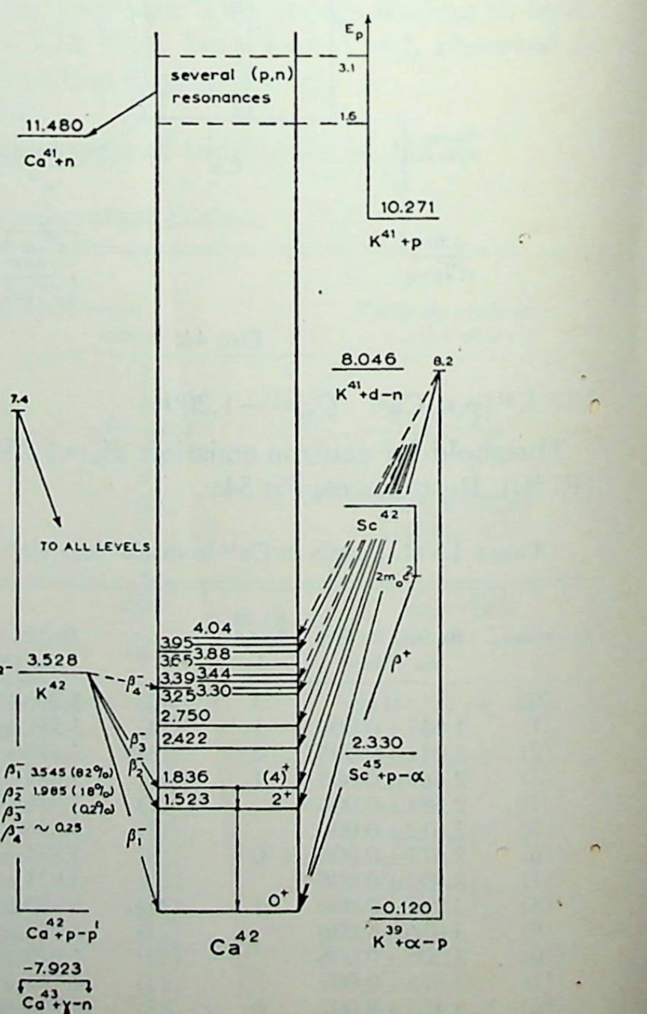


FIG. 46.

III. $K^{42}(\beta^-)Ca^{42}$ See K^{42}

IV. $Ca^{42}(p,p')Ca^{42}$

Magnetic analysis, at 90° and 120° and at several bombarding energies between 6.5 and 7.0 Mev, gives levels in Ca^{42} at 1.523 ± 0.004, 1.836 ± 0.004, 2.422 ± 0.005, 2.750 ± 0.005, 3.250 ± 0.006, 3.297 ± 0.006, 3.389 ± 0.006, 3.442 ± 0.006, 3.651 ± 0.006, 3.883 ± 0.007, 3.949 ± 0.007, and 4.043 ± 0.007 Mev (Br 56f).

V. $Sc^{42}(\beta^+)Ca^{42}$

A 0.62 ± 0.05 sec β^+ activity observed from bombardment of natural potassium with 18-Mev α particles. About same half-life estimated for a superallowed $0^+\rightarrow 0^+$ transition between $T=1$ states (Mo 55c). Also, Mo 54.

VI. Following reactions leading to Ca^{42} have not been observed: $K^{41}(p,\gamma)Ca^{42}$ ($Q_m=10.271$), $K^{41}(d,n)Ca^{42}$ ($Q_m=8.046$), $Ca^{43}(\gamma,n)Ca^{42}$ ($Q_m=-7.923$), and $Sc^{45}(p,\alpha)Ca^{42}$ ($Q_m=2.330$).

GENERAL REMARKS

Computation of excitation energy of first level in Ca^{42}, Sc 54a, Th 56.

K^{43}

(Fig. 47)

I. $K^{43}(\beta^-)Ca^{43}$ $Q_m=1.84$

Half-life = 22.4 hr (Ov 49), 21.5 hr (Ru 52a) and 22.0 hr (Li 54a).

The β^- decay is complex; β^- and γ-ray energies and intensities in Table LIX.

Shape of β_1 spectrum is unique once-forbidden ($J=2$, yes) with log $ft=8.7$ and $\log(W_0^2-1)\,ft=10.0$ (Li 54a). All γ rays reported in Be 57 can be explained as transitions between Ca^{43} states at 0, 371, 591, 985, and 1373 kev. The β_5^- branch given in Li 54a must then be regarded as nonexistent. This interpretation entails branching ratios of 6, 76, and 16% for β^- transitions to 593-, 991-, and 1394-kev levels respectively, corresponding to log $ft=7.4$, 5.6, and 5.3.

Conclusions regarding spins and parities in Fig. 48.

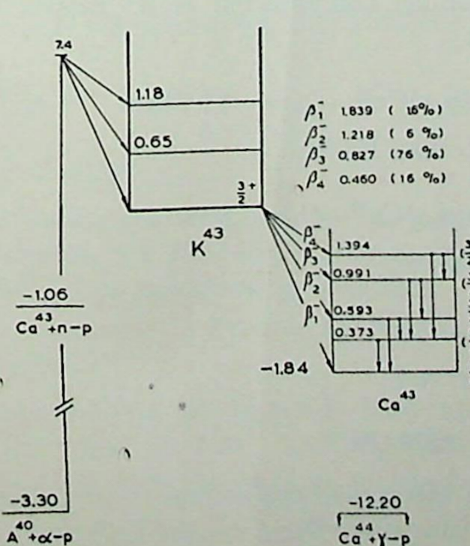


FIG. 47.

TABLE LIX. $K^{43}(\beta^-)Ca^{43}$ decay.

Reference:	Li 54a		Be 57			
Method:	magn. spectrom.		magn. spectrom.		scint. spectrom.	
	E_β, E_γ (kev)	rel. int.	E_γ (kev)	rel. int.	E_γ (kev)	rel. int.
β_1^-	1839 ± 30	1.6%				
β_2^-	1218 ± 25	5.4%				
β_3^-	827 ± 20	83.1%				
β_4^-	460 ± 20	5.4%				
β_5^-	243 ± 30	4.5%				
γ_1	219 ± 4 [a]	1			220 ± 5	<5
γ_2	369 ± 3 [a]	67	371 ± 2 [a,d]	100	375 ± 5	100
γ_3			388 ± 2 [a,d]	14		
γ_4	393 ± 4 [a]	6	394 ± 2 [a]	16		
γ_5			591 ± 2 [a,d]	23		
γ_6	627 ± 6 [a,b]	100	614 ± 4 [a]	91	615 ± 5	91
γ_7	1000 ± 20 [a,c]	4			1005 ± 10	3.5

[a] Measured from external conversion spectrum.
[b] Also internal conversion line observed with $\alpha_K\sim2\times10^{-4}$.
[c] Also observed by scintillation spectrometer (Li 54a).
[d] From β-γ and γ-γ coincidence measurements these transitions lead to Ca^{43} ground state (Be 57).

II. $A^{40}(\alpha,p)K^{43}$ $Q_m=-3.30$

From nuclear emulsion measurements, at $E_\alpha=7.4$ Mev and $\theta=90°$, ground-state Q value $=-3.36\pm0.03$ Mev, and levels in K^{43} at 0.65 ± 0.04, and 1.18 ± 0.07 Mev (Sc 56b).

III. $Ca^{43}(n,p)K^{43}$ $Q_m=-1.06$

Not observed.

IV. $Ca^{44}(\gamma,p)K^{43}$ $Q_m=-12.20$

Not observed.

Ca^{43}

(Fig. 48)

I. $A^{40}(\alpha,n)Ca^{43}$ $Q_m=-2.25$

Cross section, Sc 56b.
Resonances, Ca^{44}.

II. $Ca^{42}(n,\gamma)Ca^{43}$ $Q_m=7.923$

Cross section, Hu 55c.

The γ rays F and F' (Table LVI) from thermal-neutron capture in natural calcium assigned to $Ca^{42}(n,\gamma)Ca^{43}$ reaction (Ad 56).

III. $Ca^{42}(d,p)Ca^{43}$ $Q_m=5.708$

Many Ca^{43} levels observed from $Ca^{43}(p,p')Ca^{43}$ also found by magnetic analysis of this reaction (Table LX). Ground-state Q value $=5.711\pm0.010$ Mev (Br 56f, Br 57b). Results from angular distribution measurements obtained by magnetic analysis at $E_d=7.0$ Mev in Table LX (Bo 55b, Bo 57). Theoretical interpretation of measured angular distributions, Fr 55, Ra 56.

TABLE LX. Levels in Ca^{43}, and corresponding $Ca^{42}(d,p)Ca^{43}$ relative intensities and l_n values (Br 56f, Br 57b, Bo 55b, Bo 57).

Level	E_x (Ca^{43}) (Mev) [d]	Relative yield at maximum at $E_d=7.0$ Mev	l_n
(0)	0	100	3
(1)	0.373 ± 0.003 [a, b]	<7	
(2)	0.593 ± 0.003 [a, b]	36	1
(3)	0.991 ± 0.003 [a, b]	14	2
(4)	1.394 ± 0.004 [a, b]	<4	
(5)	1.678 ± 0.004 [a, b]	10	
(6)	1.904 ± 0.004 [a, b]	<7	
(7)	1.932 ± 0.004 [a, b]	<7	
(8)	1.957 ± 0.004 [a, b]	100 (at 10°)	0
(9)	1.985 ± 0.005 [a]	<23	
(10)	2.048 ± 0.005 [a, b]	750	1
(11)	2.069 ± 0.005 [b]	<270	
(12)	2.095 ± 0.005 [b]	<29	
(13)	(2.107 ± 0.005) [b]	<23	
(14)	2.225 ± 0.005 [a, b]	<23	
(15)	2.250 ± 0.005 [a, b]	<23	
(16)	$2.273+0.005$ [b]	<14	
(17)	2.409 ± 0.005 [a, b]	<11	
(17′)	(2.53) [c]	13	1
(18)	2.607 ± 0.005 [a, b]	84	1
(19)	2.673 ± 0.005 [b]	<14	
(20)	2.696 ± 0.005 [b]	<14	
(21)	2.753 ± 0.005 [b]	<17	
(22)	2.844 ± 0.005 [a, b]	<17	
(23)	2.880 ± 0.005 [a, b]	61	1
(24)	2.947 ± 0.005 [a, b]	61	1
(25)	3.027 ± 0.006 [b]	<11	
(26)	3.047 ± 0.006 [a, b]	<11	
(27)	3.074 ± 0.006 [b]	<11	
(28)	3.094 ± 0.006 [a, b]	<7	
(29)	3.194 ± 0.006 [b]	<27	
(30)	3.279 ± 0.006 [b]	<22	
(31)	3.293 ± 0.006 [b]	53	
(32)	3.369 ± 0.006 [b]	<14	
(33)	3.398 ± 0.006 [b]	<31	
(34)	3.419 ± 0.006 [a, b]	<31	
(35)	3.584 [c]	64	1

[a] Observed from $Ca^{42}(d,p)Ca^{43}$ reaction at several deuteron energies between 2.9 and 7.0 Mev and at $\theta=90°$ (Br 56f, Br 57b).
[b] Observed from $Ca^{43}(p,p')Ca^{43}$ reaction at $E_p=6.5$ Mev and $\theta=90°$, and at $E_p=7.0$ Mev and $\theta=130°$ (Br 56f, Br 57b).
[c] Additional level observed from $Ca^{42}(d,p)Ca^{43}$ reaction (Bo 57).
[d] Listed energies are weighted averages (where possible) of results obtained from $Ca^{42}(d,p)Ca^{43}$ and $Ca^{43}(p,p')Ca^{43}$ reactions.

IV. $K^{43}(\beta^-)Ca^{43}$ See K^{43}

V. $Ca^{43}(p,p')Ca^{43}$

Levels observed by magnetic analysis at $E_d=6.5$ and 7.0 Mev, and at $\theta=90°$ and 130° in Table LX (Br 56f, Br 57b).

TABLE LXI. Positron decay of Sc^{43}.

Reference	Method	E_{β_1+} (Mev)	E_{β_2+} (Mev)	$E\gamma$ (Mev)
Ha 52d	magn. spectrom.	1.18 ± 0.02 (72%)	0.77 ± 0.04 (28%)	0.375 ± 0.002
Nu 53	scint. spectrom.			0.375 ± 0.004 $(25\pm2)\%$
Li 54b	magn. spectrom.	1.20 ± 0.01 (82%)	0.82 ± 0.02 [a] (18%)	0.369 ± 0.005 (16%)

[a] Also reported in Li 54 are a β^+ branch of 0.39 ± 0.03 Mev endpoint (4%), and γ rays of 0.25 ± 0.01 Mev (1%), 0.627 ± 0.005 Mev (4%), and 0.84 ± 0.02 Mev (weak). Existence of last two γ rays not substantiated in later work (Va 57b).

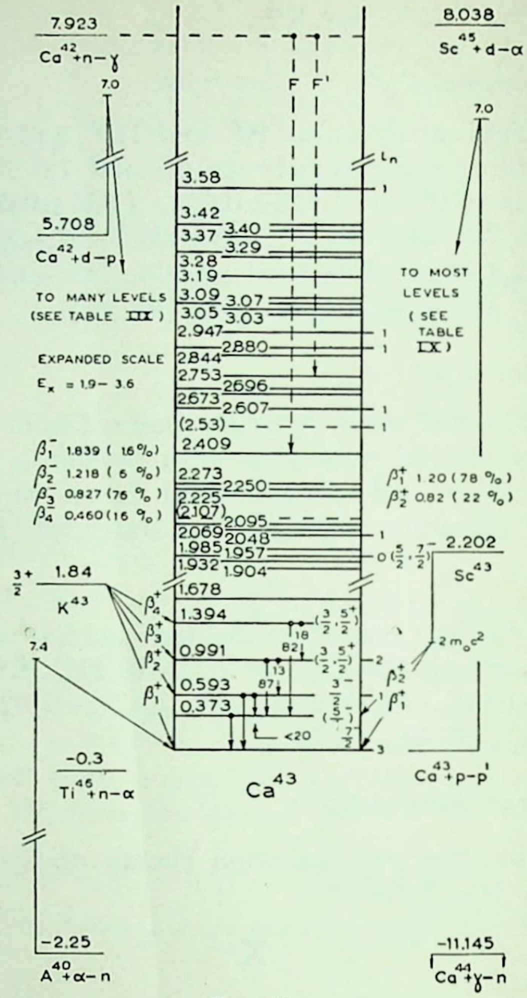


FIG. 48.

VI. $Sc^{43}(\beta^+)Ca^{43}$ $Q_m=2.202$

Most accurate measurement of half-life, 3.92 ± 0.02 hr (Hi 45), in reasonable agreement with other measurements (En 54a, Li 54).

The β^+ decay is complex; β^+ spectrum end points, relative intensities, and energies of γ rays in Table LXI. Both β^+ transitions are allowed: log $ft=5.1$ for the 1.20 Mev branch and log $ft=4.7$ for the 0.82-Mev branch.

Spin of Ca^{43} measured as $J=7/2$ by magnetic resonance (Je 53) and from hyperfine spectra (Ke 54). For jj-coupling computation of β^+ matrix elements, Gr 56c.

VII. $Ca^{44}(\gamma,n)Ca^{43}$ $Q_m=-11.145$

Cross section, En 54a.

VIII. $Sc^{45}(d,\alpha)Ca^{43}$ $Q_m=8.038$

Not observed.

IX. $Ti^{46}(n,\alpha)Ca^{43}$ $Q_m=-0.3$

Not observed.

GENERAL REMARKS

Detailed theoretical predictions of properties of Ca^{43} levels, as computed from known Ca^{41} and Ca^{42} level schemes, Le 55.

K^{44}

(not illustrated)

I. $K^{44}(\beta^-)Ca^{44}$ $Q_m=6.1$

Half-life, 18 ± 1 min (Wa 37), 20 min (An 54), and 22.0 ± 0.5 min (Co 54).

The β^- decay is complex. By scintillation spectrometer, β^- branches observed with end points of 4.9 and 1.5 Mev, and γ rays of 1.13, 2.07, 2.48, and (3.6) Mev, together with some unresolved γ rays of $E_\gamma<0.5$ Mev (Co 54). Log ft of two β^- branches would be at least 7.0 and 4.7, respectively. Probable that 4.9-Mev β^- branch proceeds to first excited state in Ca^{44}.

II. $Ca^{44}(n,p)K^{44}$ $Q_m=-5.3$

Observed, En 54a, Co 54.

III. $Ca^{46}(d,\alpha)K^{44}$

Not observed.

Ca^{44}

(Fig. 49)

I. $A^{40}(\alpha,n)Ca^{43}$ $E_b=8.90$ $Q_m=-2.25$

Resonances observed with Th C′α particles, En 54a.

II. $K^{41}(\alpha,p)Ca^{44}$ $Q_m=1.054$

Bombardment of enriched targets with 7.8-Mev α particles and range analysis, gives ground-state Q value$=0.98\pm0.10$ Mev, and levels in Ca^{44} at 1.13, 1.92, 2.28, 2.58, 2.97, and 3.17 Mev, all ±0.05 Mev (Sc 55).

III. $Ca^{43}(d,p)Ca^{44}$ $Q_m=8.920$

At $E_d=4.2$ Mev, using enriched targets and range analysis, ground-state $Q=9.07\pm0.07$ Mev, with Ca^{43} levels at 1.15 and 2.28 Mev (Sc 55). Magnetic analysis at $E_d=6.0$ Mev and $\theta=90°$, proton groups observed leading to seven lowest states in Ca^{44} from which ground-state $Q=8.913\pm0.014$ Mev (Br 56f).

IV. $K^{44}(\beta^-)Ca^{44}$ See K^{44}

V. $Ca^{44}(n,n'\gamma)Ca^{44}$

From inelastic scattering of 3.9-Mev neutrons on natural calcium, a 1.152 ± 0.020-Mev γ ray observed with a scintillation spectrometer (Da 56c).

VI. $Ca^{44}(p,p')Ca^{44}$

Magnetic analysis at $E_p=6.5$ Mev and $\theta=90°$, $E_p=7.0$ Mev and $\theta=120°$, and $E_p=7.4$ Mev and $\theta=90°$, levels in Ca^{44} observed at 1.156 ± 0.004, 1.883 ± 0.004, 2.284 ± 0.005, 2.655 ± 0.005, 3.044 ± 0.005, 3.297 ± 0.006, 3.305 ± 0.006, 3.354 ± 0.006, 3.581 ± 0.006, 3.656 ± 0.006, and 3.671 ± 0.006 Mev (Br 56f).

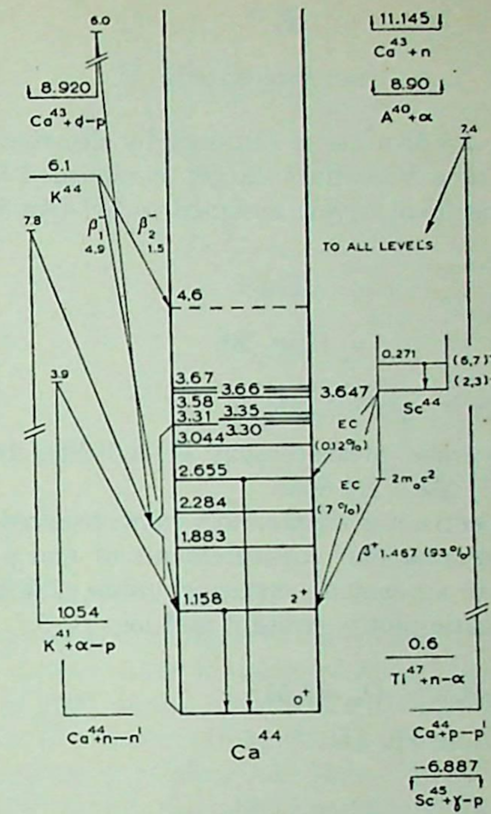


FIG. 49.

VII. (a) $Sc^{44}(\beta^+)Ca^{44}$ $Q_m=3.647$

Most accurate measurement of half-life, 3.92 ± 0.03 hr (Hi 45). Other measurements (En 54a) in reasonable agreement.

The β^+ end point, measured by magnetic spectrometer, 1.463 ± 0.005 Mev (Br 50b), and 1.471 ± 0.005 Mev (Bl 55), yielding log $ft=5.3$. Kurie plot is straight. For jj-coupling calculation of the β^+ matrix element, Gr 56c. The β^+ decay is followed by a γ ray with $E_\gamma=1.159\pm0.003$ Mev. Its conversion coefficient is $(6.3\pm0.3)\times10^{-5}$ establishing this transition as E2 (Bl 55). Electron capture also occurs with an intensity measured as $6.8\pm1.5\%$ per disintegration (Bl 55, also La 54a). Theoretical value for an allowed transition is 5.8% (Zw 54).

A weak 2.54 ± 0.03-Mev γ ray also observed with an intensity of $0.52\pm0.02\%$ per disintegration. No 1.38-Mev γ ray found either in single spectrum or in spectrum coincident with 1.16-Mev γ ray (intensity $<0.5\%$ per disintegration) (Bl 55).

(b) $Sc^{44m}(\gamma)Sc^{44}$.

Most accurate measurement of half-life, 58.6 ± 0.7 hr (Hi 45, Ba 51). Decay to Sc^{44} ground state through a γ ray of energy 269.3 ± 1 kev (Sm 42) or 271.3 ± 0.7 kev (Br 50b). Total conversion coefficient 0.139 ± 0.003 in agreement with that expected for an E4 transition (Bl 55).

Also, En 54a.

VIII. Following reactions leading to Ca^{44} have not been observed: $Ca^{43}(n,\gamma)Ca^{44}$ $(Q_m=11.145)$, $Sc^{45}(\gamma,p)Ca^{44}$ $(Q_m=-6.887)$, and $Ti^{47}(n,\alpha)Ca^{44}$ $(Q_m=0.6)$.

K^{45}.

(not illustrated)

A 34 min $A=45$ activity obtained by electromagnetic separation of a vanadium target bombarded by 187-Mev protons. Tentatively assigned to K^{45} (An 54).

Ca^{45}

(Fig. 50)

I. $Ca^{45}(\beta^-)Sc^{45}$ $Q_m=0.256$

Most accurate measurement of half-life, 163.5±4 days (De 53). Also En 54a.

The β^- spectrum is simple, no γ rays observed. Three magnetic spectrometer measurements of end point in good mutual agreement, average value, 256±2 kev (En 54a). Kurie plot is straight and log $ft=5.7$.

II. $Ca^{44}(n,\gamma)Ca^{45}$ $Q_m=7.413$

Cross section, Hu 55c.

III. $Ca^{44}(d,p)Ca^{45}$ $Q_m=5.188$

By magnetic analysis, at several deuteron energies between 2.9 and 7.0 Mev and $\theta=90°$, ground-state Q value=5.188±0.010 Mev (Br 56f). Levels in Ca^{45} (Br 56f), and angular distribution measurements at $E_d=7.0$ Mev, also by magnetic analysis (Co 57), in Table LXII.

Large intensity of ground-state transition makes it probable that Ca^{45} ground state is a $f_{7/2}$ single-neutron state.

Theoretical remarks on the angular distributions, Fr 55, Ra 56.

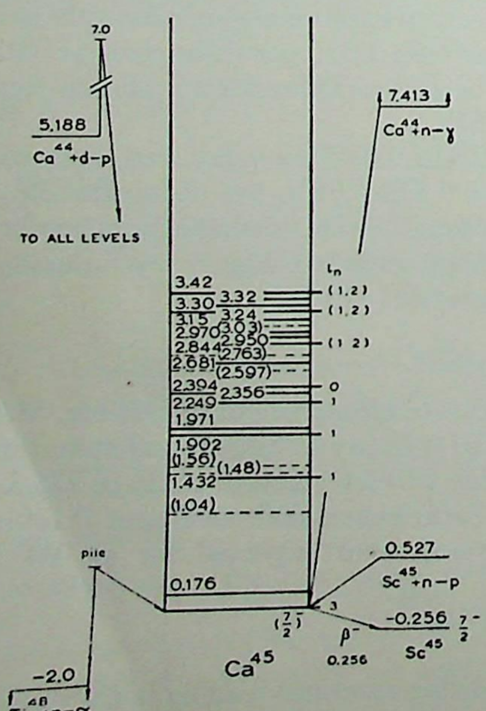


FIG. 50.

TABLE LXII. Levels in Ca^{45} from $Ca^{44}(d,p)Ca^{45}$.

Level	E_x (Ca^{45}) (Mev) [a]	Angle of maximum yield [b]	Relative yield at maximum [b]	l_n [b]
(0)	0	40°	100	3
(1)	0.176±0.003	50°	3.0	...
(2)	(1.036±0.010)		<11	
(3)	1.432±0.004	20°	200	1
(4)	(1.475±0.006)		<7	
(5)	(1.557±0.010)		<6	
(6)	1.902±0.004	20°	115	1
(7)	1.971±0.004	50°	3.1	...
(8)	2.249±0.005	15°	190	1
(9)	2.356±0.005	30°	1.7	...
(10)	2.394±0.005	<7½°	190 (at 7½°)	0
(11)	(2.597±0.005)		<4	
(12)	2.681±0.005	20°	3.9	...
(13)	(2.763±0.005)		<8	
(14)	2.844±0.005	20°	212	(1,2)
(15)	2.950±0.005		<5	
(16)	2.970±0.005	10°	19.5	...
(17)	(3.032±0.006)			
(18)	3.148±0.006		<5	
(19)	3.244±0.006	20°	104	(1,2)
(20)	3.296±0.006		<4	
(21)	3.319±0.006	40°	18.9	...
(22)	3.419±0.006	15°	430	(1,2)

[a] From Br 56f.
[b] From Co 57.

IV. $Sc^{45}(n,p)Ca^{45}$ $Q_m=0.527$

Observed, En 54a.

V. $Ca^{46}(\gamma,n)Ca^{45}$

Not observed.

VI. $Ti^{48}(n,\alpha)Ca^{45}$ $Q_m=-2.0$

Cross section, En 54a.

Ca^{46}

(not illustrated)

No reactions leading to Ca^{46} have been observed. Possibilities would be $Ca^{44}(t,p)Ca^{46}$ or $Ti^{49}(n,\alpha)Ca^{46}$. Nor has Ca^{46} been concentrated sufficiently to make a study of inelastic scattering feasible.

Ca^{47}

(not illustrated)

I. $Ca^{47}(\beta^-)Sc^{47}$ $Q_m=2.0$

Half-life average from several measurements (En 54a, Ly 55, Li 56, see also Wa 55) 4.7±0.2 days.

The β^- decay is complex (Table LXIII; also En 54a). Log ft values for β_1 and β_2 are 8.5 and 6.0 respectively. The first is unexpectedly large for a presumably allowed transition (also Gr 56c).

II. $Ca^{46}(n,\gamma)Ca^{47}$

Observed, En 54a.

TABLE LXIII. $Ca^{47}(\beta^-)Sc^{47}$.

Reference: Method:	Co 53 [a] magn. spectrom.	Ma 53 magn. spectrom.	Ly 55 Al abs. +scint. spectrom.	Li 56 magn. spectrom. +scint. spectrom.
E_{β_1} (Mev)	1.4 ±0.1 (40%)	2.060±0.020 [b] (19%)	1.93±0.02 (24±6)%	1.940±0.020 [b] (17%)
E_{β_2} (Mev)	0.46 ±0.02 (60%)	0.685±0.006 (81%)	0.70±0.02 (76±6)%	0.660±0.010 [b] (83%)
E_{γ_1} (Mev)	1.303±0.040		1.29 (14.2±1.2)	1.31 ±0.02 [c] (13±2)
E_{γ_2} (Mev)	0.800±0.025		0.812 (1±0.1)	0.83 ±0.02 [c] (1±0.1)
E_{γ_3} (Mev)	0.495±0.015		0.500 (1)	0.48 ±0.02 [c] (1)

[a] Additional γ rays of 234.0 and 149.5 kev reported in Co 53, not observed by others.
[b] Kurie plot is straight.
[c] Coincidences observed between γ_2 and γ_3; neither γ_2 nor γ_3 coincident with γ_1.

III. Following reactions leading to Ca^{47} have not been reported: $Ca^{46}(d,p)Ca^{47}$, $Ca^{48}(\gamma,n)Ca^{47}$ ($Q_m=-11.0$), $Ti^{50}(n,\alpha)Ca^{47}$ ($Q_m=-3.5$).

Ca^{48}

(not illustrated)

I. $Ca^{48}(\beta^-)Sc^{48}$ $Q_m=0.2$

Half-life$>2\times10^{16}$ yr (Jo 52). Also, En 54a.

II. $Ca^{48}(\beta^-\beta^-)Ti^{48}$ $Q_m=4.2$

Double β^- decay of Ca^{48} with half-life of $(1.6\pm0.7)\times10^{17}$ yr and total kinetic energy of two electrons of 4.1 ± 0.3 Mev reported (Mc 55). However, lower limit for half-life given as 2×10^{18} yr (Aw 56).

III. $Ca^{48}(p,p')Ca^{48}$

From magnetic analysis at $E_p=6.5$, 7.0, and 7.4 Mev and at several angles of observation, levels in Ca^{48} observed at 3.825 ± 0.006 and 4.499 ± 0.007 Mev (Br 56f). Search for de-excitation of the first level by internal pair formation unsuccessful (Kr 57).

Ca^{49}

(Fig. 51)

I. $Ca^{49}(\beta^-)Sc^{49}$ $Q_m=5.1$

Half-life averaged from two measurements (Ke 56, Ma 56a) is 8.83 ± 0.14 min.

The β^- decay is complex (Table LXIV). Beta transitions with log ft values of 4.9, 4.6, and 4.5 occur to Sc^{49} levels at 3.09, 4.05, and 4.68 Mev respectively, which are de-excited by γ-ray transitions directly to Sc^{49} ground state.

II. $Ca^{48}(n,\gamma)Ca^{49}$ $Q_m=5.1$

Cross section, En 54a.

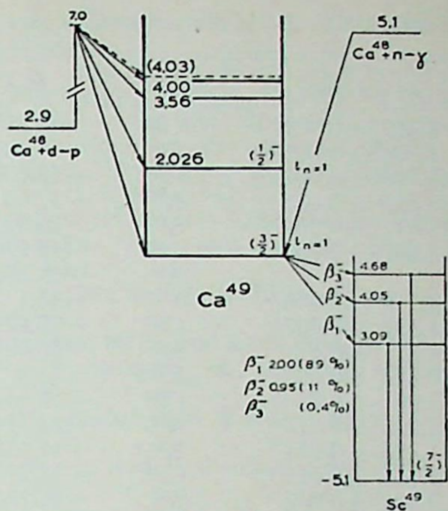


FIG. 51.

III. $Ca^{48}(d,p)Ca^{49}$ $Q_m=2.9$

From magnetic analysis at $E_d=6.5$ and 7.0 Mev and $\theta=90°$, ground-state Q value$=2.916\pm0.006$ Mev, and levels in Ca^{49} found at 2.026 ± 0.005, 3.589 ± 0.006, 4.004 ± 0.007, and (4.026 ± 0.010) Mev (Br 56f). By Al absorption, $Q=2.80\pm0.30$ Mev (Wa 54).

Angular distribution measurements at $E_d=6.5$ Mev, also by magnetic analysis, show that both ground state and first level are p states ($l_n=1$) (Bu 54).

Atomic Mass Excesses

Atomic mass excesses used for computation of Q_m values given in the present compilation in Table LXV.

Masses of most nuclides with $A>32$ are from En 57. They were computed from Q-value chains connecting them to S^{32}, for which nuclide the accurate mass-spectroscopic value was used (Qu 56a). Masses of nuclides in the $16<A<32$ region are obtained from those computed by Wapstra (Wa 55a) from Q-value measurements. However, Wapstra finds a S^{32} mass which is smaller by 41.7 kev than the mass-spectroscopic value. To obtain a smooth fit to the $A>32$ region an appropriate correction was applied to the $A<32$ masses given by Wapstra. Of course, this correction is small for nuclides near O^{16}, and relatively large for nuclides near S^{32}. In principle the correction can be obtained from a lengthy and tedious least-squares analysis. This was avoided by the use of a resistor net-

TABLE LXIV. $Ca^{49}(\beta^-)Sc^{49}$.

Reference: Method:	Ke 56 [a] scint. spectrom.	Ma 56a [b] scint. spectrom.
E_{β_1} (Mev)	1.95±0.05 (88%)	2.12±0.10
E_{β_2} (Mev)	0.95±0.15 (12%)	~1.0
E_{γ_1} (Mev)	3.10±0.03 (90±2%)	3.07±0.05 (89%)
E_{γ_2} (Mev)	4.05±0.05 (10±2%)	4.04±0.06 (10%)
E_{γ_3} (Mev)	4.68±0.05 (0.38±0.1%)	4.7 ±0.1 (0.8%)

[a] Intensity of a possible 0.95-Mev γ ray is $<2\%$. No γ-γ coincidences are observed.
[b] Intensity of ground-state β^- transition is $<1\%$. Intensity of a possible 0.97 Mev γ ray is $<3\%$.

TABLE LXV. Table of atomic mass excesses.

Nuclide	M-A (Mev)	Nuclide	M-A (Mev)
n^{1}	8.3675±0.0014 [a]	Cl^{34}	−14.35 ±0.03 [n]
H^{1}	7.5845±0.0014 [a]	S^{35}	−18.489 ±0.008 [d]
H^{2}	13.7256±0.0026 [a]	Cl^{35}	−18.657 ±0.008 [d]
H^{3}	15.835 ±0.005 [a]	A^{35}	−12.68 ±0.04 [n]
He^{3}	15.817 ±0.005 [a]	S^{36}	−20.04 ±0.06 [o]
He^{4}	3.6073±0.0024 [a]	Cl^{36}	−18.870 ±0.018 [d]
		A^{36}	−19.584 ±0.018 [d]
F^{19}	4.147 ±0.006 [b]	S^{37}	−16.1 ±0.1 [p]
Ne^{20}	−1.137 ±0.007 [b]	Cl^{37}	−20.824 ±0.014 [d]
Na^{20}	14.2 ±0.5 [b]	A^{37}	−20.009 ±0.015 [d]
Ne^{21}	0.473 ±0.008 [b]	K^{37}	−13.9 ±0.1 [n]
Na^{21}	3.99 ±0.03 [b]	Cl^{38}	−18.564 ±0.017 [d]
Ne^{22}	−1.525 ±0.013 [b]	A^{38}	−23.474 ±0.032 [d]
Na^{22}	1.315 ±0.014 [b]	K^{38} [z]	−17.558 ±0.026 [d]
Ne^{23}	1.651 ±0.015 [b]	Cl^{39}	−18.274 ±0.026 [n]
Na^{23}	−2.734 ±0.007 [b]	A^{39}	−21.714 ±0.021 [d]
Mg^{23}	1.325 ±0.012 [f]	K^{39}	−22.279 ±0.020 [d]
Ne^{24}	1.13 ±0.04 [g]	Ca^{39}	−15.5 ±0.2 [q]
Na^{24}	−1.323 ±0.008 [b]	Cl^{40}	−15.7 ±1.0 [n]
Mg^{24}	−6.837 ±0.008 [b]	A^{40}	−23.19 ±0.03 [r]
Al^{24}	7.2 ±0.3 [b]	K^{40}	−21.707 ±0.021 [d]
Na^{25}	−1.80 ±0.20 [h]	Ca^{40}	−23.026±0.024 [d]
Mg^{25}	−5.802±0.009 [b]	Sc^{40}	−9.1 ±0.4 [s]
Al^{25}	−1.54 ±0.02 [i]	A^{41}	−20.925±0.035 [d]
Mg^{26}	−8.553±0.010 [b]	K^{41}	−23.449±0.028 [d]
Al^{26}	−4.537±0.018 [j]	Ca^{41}	−23.023±0.026 [d]
Mg^{27}	−6.625±0.010 [b]	Sc^{41}	−17.1 ±0.1 [t]
Al^{27}	−9.219±0.007 [b]	K^{42}	−22.608±0.033 [d]
Si^{27}	−4.414±0.009 [k]	Ca^{42}	−26.136±0.021 [d]
Mg^{28}	−6.765±0.022 [b]	K^{43}	−23.86 ±0.03 [n]
Al^{28}	−8.575±0.009 [b]	Ca^{43}	−25.702±0.023 [d]
Si^{28}	−13.223±0.009 [b]	Sc^{43}	−23.500±0.030 [d]
P^{28}	0.6 ±0.3 [b]	K^{44}	−22.4 ±0.3 [n]
Al^{29}	−9.35 ±0.10 [b]	Ca^{44}	−28.480±0.026 [d]
Si^{29}	−13.330±0.009 [b]	Sc^{44}	−24.833±0.027 [u]
P^{29}	−8.363±0.009 [b]	Ca^{45}	−27.527±0.028 [d]
Si^{30}	−15.572±0.010 [b]	Sc^{45}	−27.783±0.028 [d]
P^{30}	−11.308±0.012 [l]	Sc^{46}	−28.41 ±0.08 [e]
Si^{31}	−13.799±0.009 [b]	Ti^{46}	−30.77 ±0.08 [e]
P^{31}	−15.280±0.008 [b]	Ca^{47}	−28.64 ±0.08 [v]
S^{31}	−9.83 ±0.07 [b]	Sc^{47}	−30.59 ±0.08 [w]
Si^{32}	−14.73 ±0.05 [b]	Ti^{47}	−31.19 ±0.08 [e]
P^{32}	−14.833 ±0.004 [b]	Ca^{48}	−30.26 ±0.12 [x]
S^{32}	−16.5373±0.0008 [c]	Sc^{48}	−30.35 ±0.09 [e]
Cl^{32}	−3.5 ±0.4 [b]	Ti^{48}	−34.34 ±0.08 [e]
P^{33}	−16.567 ±0.010 [n]	Ca^{49}	−27.04 ±0.12 [y]
S^{33}	−16.816 ±0.008 [d]	Sc^{49}	−32.09 ±0.10 [e]
Cl^{33}	−11.238 ±0.012 [m]	Ti^{49}	−34.09 ±0.08 [e]
P^{34}	−14.67 ±0.14 [n]	Ti^{50}	−36.71 ±0.08 [e]
S^{34}	−19.868 ±0.016 [d]		

[a] Wa 5Fa.
[b] Wa 55a plus correction as described in text.
[c] Qu 56a. *Note added in proof.*—A recent remeasurement [Quisenberry, Giesen, and Benson, Bull. Am. Phys. Soc. Ser. II, 2, 223 (1957)] gives −16.5396±0.0010 Mev for the S^{32} mass excess.
[d] En 57.
[e] Wa 55b.
[f] Connected to Na^{23} through a new value for the $Na^{23}(p,n)Mg^{23}$ threshold (Ki 55a).
[g] Connected to Na^{24} through the $Ne^{24}(\beta^-)Na^{24}$ decay (Dr 56).
[h] Connected to Mg^{25} through the $Na^{25}(\beta^-)Mg^{25}$ decay (Ma 55).
[i] Computed from the $Al^{25}(\beta^+)Mg^{25}$ decay (El 55), and from $Mg^{24}(p,\gamma)Al^{25}$ (Hu 55, Ag 56, Cr 56, Va 56b), which reactions yield Al^{25} masses in good mutual agreement.
[j] Values for Al^{26} mass excess computed from $Si^{28}(d,\alpha)Al^{26}$ (Br 54) and $Mg^{26}(p,n)Al^{26m}$ (Ki 55a) Q values are −4.521±0.012 and −4.558±0.014 Mev respectively, in rather serious disagreement. A weighted average has been taken, which agrees well with measured $Al^{26m}(\beta^+)Mg^{26}$ decay energy (Ka 55, El 55).
[k] Connected to Al^{27} through recent measurements of $Al^{27}(p,n)Si^{27}$ threshold (Ki 55a, Ma 55a).
[l] Connected to S^{32} through $S^{32}(d,\alpha)P^{30}$ reaction' (Pa 57). P^{30} mass thus obtained is in good agreement with less accurate values computed from $P^{30}(\beta^+)Si^{30}$ decay (Gr 56), and from $Si^{29}(p,\gamma)P^{30}$ (Va 57), $Si^{29}(d,n)P^{30}$ (Ma 52), and $P^{31}(\gamma,n)P^{30}$ (Sc 55a) reactions. Serious disagreement with Q values reported in Bu 56c for $Al^{27}(\alpha,n)P^{30}$, and in Le 56b for $S^{32}(d,\alpha)P^{30}$ reaction.
[m] Connected to S^{32} through the $S^{32}(p,\gamma)Cl^{33}$ reaction (Va 56f).
[n] Connected through beta decay to nuclides listed in En 57.
[o] From microwave measurements yielding $(S^{36}-S^{32})/(S^{36}-S^{34})=1.998320\pm30$ (Ge 56c).
[p] From $S^{37}(\beta^-)Cl^{37}$ (Bl 46, Mo 56, St 56a) and $A^{40}(n,\alpha)S^{37}$ (Be 55a) Q values in good mutual agreement.
[q] Average of masses computed from $Ca^{39}(\beta^+)K^{39}$ (Hu 54) and $Ca^{40}(\gamma,n)Ca^{39}$ (Su 53) Q values.
[r] Mass adopted for A^{40} makes the energy available for $K^{40}(EC)A^{40}$ decay leading to first A^{40} level equal to only 20 kev which is necessary to explain the long observed half-life (also Wa 55). This mass in fair agreement with that computed from $A^{40}(d,p)A^{41}$ Q value (Da 49, Gi 52). *Note added in proof.*—The adopted A^{40} mass excess is in excellent agreement with a recent precision measurement by mass synchrometer [L. G. Smith, Bull. Am. Phys. Soc. Ser. II, 2, 223 (1957)] yielding −23.1893±0.0004 Mev.
[s] Average of masses computed from $Sc^{40}(\beta^+)Ca^{40}$ and $Ca^{40}(p,n)Sc^{40}$ Q values (Gl 55).
[t] Average of masses computed from $Sc^{41}(\beta^+)Ca^{41}$ (El 41) and $Ca^{40}(d,n)Sc^{41}$ (Pl 55) Q values.
[u] Value given here slightly different from that in En 57, where $Sc^{44}(\beta^+)Ca^{44}$ decay energy reported in Bl 55 had been overlooked.
[v] Connected to Sc^{47} through $Ca^{47}(\beta^-)Sc^{47}$ decay (Co 53, Ma 53, Ly 55a, Li 56).
[w] Connected to Ti^{47} through $Sc^{47}(\beta^-)Ti^{47}$ decay (Ch 53a, Ni 55, Gr 56a).
[x] Connected to Ca^{49} through $Ca^{48}(d,p)Ca^{49}$ Q value (Br 56f).
[y] Connected to Sc^{49} through $Ca^{49}(\beta^-)Sc^{49}$ decay (Ke 56, Ma 56a).
[z] Lowest $T=0$ state.

work, to be described elsewhere. Each resistor represents a Q-value measurement. Its resistance is taken proportional to the square of the experimental error of the corresponding measurement. Resistors are soldered to jacks representing nuclides. If an external potential difference of 41.7 v is applied between the S^{32} and O^{16} jacks, the potential difference (in v) between any arbitrary jack and O^{16} gives the corresponding mass correction (in kev). The error to be assigned to any mass value is obtained by measuring the resistance between the corresponding jack and O^{16}.

REFERENCES

Aa 55 L. C. Aamodt and P. C. Fletcher, Phys. Rev. **98**, 1317 (1955).
Ad 53 Adler, Huber, and Hälg, Helv. Phys. Acta **26**, 349 (1953).
Ad 56 Adyasevich, Groshev, Demidov, and Lysenko, Atomnaja Energ. **2**, 28 (1956).
Ad 56a Adyasevich, Groshev, and Demidov, Atomnaja Energ. **2**, 40 (1956), and J. Nuclear Energy **3**, 258 (1956).
Ag 50 H. M. Agnew, Phys. Rev. **77**, 655 (1950).
Ag 56 Ager-Hanssen, Lönsjö, and Nordhagen, Phys. Rev. **101**, 1779 (1956).
Ah 48 L. H. Ahrens and R. D. Evans, Phys. Rev. **74**, 279 (1948).
Aj 52 F. Ajzenberg and T. Lauritsen, Revs. Modern Phys. **24**, 321 (1952).
Aj 55 Ajzenberg, Rubin, and Likely, Phys. Rev. **99**, 654(A) (1955) plus verbal report to Nuclear Data Group.
Al 48 D. E. Alburger, Phys. Rev. **73**, 1014 (1948).
Al 49 D. E. Alburger, Phys. Rev. **76**, 435(L) (1949).
Al 49a D. E. Alburger, Phys. Rev. **75**, 51 (1949).
Al 50 D. E. Alburger and E. M. Hafner, Revs. Modern Phys. **22**, 373 (1950).
Al 50a L. W. Alvarez, Phys. Rev. **80**, 519 (1950).
Al 50b Alburger, Hughes, and Eggler, Phys. Rev. **78**, 318(A) (1950).
Al 51 Allen, May, and Rall, Phys. Rev. **84**, 1203 (1951).
Al 55 Allen, Burcham, Chackett, Munday, and Reasbeck, Proc. Phys. Soc. (London) **A68**, 681 (1955).
Al 55a Almqvist, Clarke, and Paul, Phys. Rev. **100**, 1265(A) (1955).
Al 55b H. R. Allan and N. Sarma, Proc. Phys. Soc. (London) **A68**, 535 (1955).
Al 56 Alkhazov, Andreyev, Greenberg, and Lemberg, Physica **22**, 1129(A) (1956).
Al 56a Allen, Collinge, Hird, Maglić, and Orman, Proc. Phys. Soc. (London) **A69**, 705 (1956).
Al 57 Almqvist, Bromley, Ferguson, Gove, Litherland, and Paul (private communication).
Al 57a D. L. Allan, Proc. Phys. Soc. (London) **A70**, 195 (1957).
An 52 Anderson, Wheeler, and Watson, Phys. Rev. **87**, 897(L) (1952).
An 53 Anderson, Wheeler, and Watson, Phys. Rev. **90**, 606 (1953).

An 54 G. Andersson, Phil. Mag. **45**, 621 (1954).
An 45a Anton'eva, Bashilov, Dzhelepov, and Orlov, Izvest. Akad. Nauk. Ser. Fiz. SSSR **18**, 93 (1954).
Ar 52 Arthur, Allen, Bender, Hausman, and McDole, Phys. Rev. **88**, 1291 (1952).
Ar 53 W. Arber and P. Stähelin, Helv. Phys. Acta **26**, 584(A) (1953).
Aw 56 M. Awschalom, Phys. Rev. **101**, 1041 (1956).
Ba 40 Barkas, Creutz, Delsasso, Sutton, and White, Phys. Rev. **58**, 383(L) (1940).
Ba 51 Batzel, Miller, and Seaborg, Phys. Rev. **84**, 671 (1951).
Ba 52 Baker, Dodd, and Simmons, Phys. Rev. **85**, 1051(L) (1952).
Ba 52a K. Baskova and A. Kudriavtseva, Zhur. Eksptl. i Teoret. Fiz. **23**, 483 (1952).
Ba 53 G. A. Bartholomew and B. B. Kinsey, Can. J. Phys. **31**, 927 (1953).
Ba 54 R. E. Batzel and G. H. Coleman, Phys. Rev. **93**, 280 (1954).
Ba 55 R. Basile and C. Schuhl, Compt. rend. **240**, 2399 (1955).
Ba 55a Bartholomew, Boyd, Brown, Hawkings, Lounsbury, and Merritt, Can. J. Phys. **33**, 43 (1955).
Ba 55b G. Backenstoss and K. Goebel, Z. Naturforsch. **10a**, 920 (1955).
Ba 56 Baumann, Prosser, Read, and Krone, Phys. Rev. **104**, 376 (1956).
Be 39 H. A. Bethe and W. J. Henderson, Phys. Rev. **56**, 1060(L) (1939).
Be 47 Becker, Hanson, and Diven, Phys. Rev. **71**, 466(A) (1947).
Be 48 B. B. Benson, Phys. Rev. **73**, 7 (1948).
Be 49 Bender, Shoemaker, Kaufmann, and Bouricius, Phys. Rev. **76**, 273 (1949).
Be 50 J. R. Beyster and M. L. Wiedenbeck, Phys. Rev. **79**, 728(L) (1950).
Be 51 Beghian, Bishop, and Halban, Phys. Rev. **83**, 186(L) (1951).
Be 53 E. H. Bellamy and K. F. Smith, Phil. Mag. **44**, 33 (1953).
Be 55 Bent, Bonner, McCrary, and Ranken, Phys. Rev. **100**, 774 (1955).
Be 55a E. H. Bellamy and F. C. Flack, Phil. Mag. **46**, 341 (1955).
Be 55b Bent, Bonner, and McCrary, Phys. Rev. **98**, 1325 (1955).
Be 55c H. Berthet and J. Rossel, Helv. Phys. Acta **28**, 265 (1955).
Be 56 Beyster, Walt, and Salmi, Phys. Rev. **104**, 1319 (1956).
Be 56a Berenbaum, Towle, and Matthews, Proc. Phys. Soc. (London) **A69**, 858(L) (1956).
Be 57 Benczer-Koller, Schwartzschild, and Wu, Bull. Am. Phys. Soc. Ser. II, **2**, 23 (1957).
Bi 52 A. C. Birge, Phys. Rev. **85**, 753(A) (1952).
Bl 46 E. Bleuler and W. Zünti, Helv. Phys. Acta **19**, 137 (1946).
Bl 47 E. Bleuler and W. Zünti, Helv. Phys. Acta **20**, 195 (1947).
Bl 51 Blaser, Boehm, Marmier, and Scherrer, Helv. Phys. Acta **24**, 465 (1951).
Bl 55 J. W. Blue and E. Bleuler, Phys. Rev. **100**, 1324 (1955).
Bl 55a E. Bleuler and D. J. Tendam, Phys. Rev. **99**, 1652(A) (1955).
Bl 56 R. C. Block and H. W. Newson, Bull. Am. Phys. Soc. Ser. II, **1**, 55 (1956).
Bl 56a Bleuler, Gailar, Seidlitz, and Tendam, Physica **22**, 1127(A) (1956).
Bo 50 K. Boyer, M.I.T., Laboratory for Nuclear Science and Engineering Progress Report, July 1, 1950.
Bo 51 F. J. Boley and D. J. Zaffarano, Phys. Rev. **84**, 1059(L) (1951).
Bo 53 F. J. Boley, Iowa State Coll. J. Sci. **27**, 129 (1953).
Bo 55 M. Bolsterli and E. Feenberg, Phys. Rev. **97**, 736 (1955).
Bo 55a Boyd, Brown, and Lounsbury, Can. J. Phys. **33**, 35 (1955).
Bo 55b Bockelman, Braams, Buechner, and Guthe, Phys. Rev. **99**, 655(A) (1955).
Bo 55c Bockelman, Cobb, and Guthe, M.I.T., Laboratory for Nuclear Science Progress Report, May 31 (1955).
Bo 56 C. K. Bockelman, Bull. Am. Phys. Soc. Ser. II, **1**, 223 (1956).
Bo 56a C. K. Bockelman, M.I.T., Laboratory for Nuclear Science Progress Report, May 31 (1956).
Bo 57 Bockelman, Braams, Browne, Buechner, Sharp, and Sperduto, Phys. Rev. **107**, 176 (1957).
Bo 57a Bostrom, Moore, and Morgan, Bull. Am. Phys. Soc. Ser. II, **2**, 104 (1957).
Br 47 Broström, Huus, and Koch, Nature **160**, 498(L) (1947).
Br 47a Broström, Huus, and Tangen, Phys. Rev. **71**, 661 (1947).
Br 48 H. Bradner and J. D. Gow, Phys. Rev. **74**, 1559(A) (1948).
Br 48a Broström, Huus, and Koch, Nature **162**, 695(L) (1948).
Br 49 Brolley, Sampson, and Mitchell, Phys. Rev. **76**, 624 (1949).
Br 50 Brosi, Zeldes, and Ketelle, Phys. Rev. **79**, 902(L) (1950).
Br 50a H. Brown and V. Perez-Mendez, Phys. Rev. **78**, 649 (1950).
Br 50b J. A. Bruner and L. M. Langer, Phys. Rev. **79**, 606 (1950).
Br 51 Broström, Madsen, and Madsen, Phys. Rev. **83**, 1265(L) (1951).
Br 51a Brown, Hanna, and Yaffe, Phys. Rev. **84**, 1243(L) (1951).
Br 53 R. Braams and C. L. Smith, Phys. Rev. **90**, 995(L) (1953).
Br 53a Brown, Yaffe, and Hanna, Proc. Roy. Soc. (London) **220A**, 203 (1953).
Br 54 C. P. Browne, Phys. Rev. **95**, 860(L) (1954).
Br 54a Breckon, Henrikson, Martin, and Foster, Can. J. Phys. **32**, 223 (1954).
Br 54b Browne, Zimmerman, and Buechner, Phys. Rev. **96**, 725 (1954).
Br 54c Bretscher, Alderman, Elwyn, and Shull, Phys. Rev. **96**, 103 (1954).
Br 55 C. P. Browne and W. C. Cobb, Phys. Rev. **99**, 644(A) (1955).
Br 55a C. P. Browne, M.I.T., Laboratory for Nuclear Science Progress Report, May 31 (1955).
Br 55b R. J. Breen and M. R. Hertz, Phys. Rev. **98**, 599 (1955).
Br 56 Bromley, Gove, Litherland, Paul, and Almqvist, Bull. Am. Phys. Soc. Ser. II, **1**, 30 (1956).
Br 56a Broude, Green, Singh, and Willmott, Phys. Rev. **101**, 1052 (1956).
Br 56b C. M. Braams, Phys. Rev. **101**, 1764 (1956).
Br 56c Brugger, Evans, Joki, and Shankland, Phys. Rev. **104**, 1054 (1956).
Br 56d C. P. Browne, Bull. Am. Phys. Soc. Ser. II, **1**, 212 (1956) and M.I.T., Laboratory for Nuclear Science Progress Report, May 31 (1956).
Br 56e T. H. Braid, Phys. Rev. **102**, 1109 (1956).
Br 56f C. M. Braams, thesis Utrecht (1956).
Br 56g C. M. Braams, Phys. Rev. **103**, 1310 (1956).
Br 56h Broude, Green, Singh, and Willmott (private communication to Nuclear Data Group).
Br 56i Broude, Green, Willmott, and Singh, Physica **22**, 1139(A) (1956), and private communication.
Br 57 Bromley, Gove, Paul, Litherland, and Almqvist, Can. J. Phys. **35**, 1042 (1957).
Br 57a Bromley, Gove, Litherland, and Almqvist (private communication).
Br 57b C. M. Braams, Phys. Rev. **105**, 1023 (1957).
Br 57c Bromley, Gove, and Litherland, Can. J. Phys. **35**, 1057 (1957).
Bu 53 P. R. J. Burch, Nature **172**, 361(L) (1953).
Bu 53a Buechner, Sperduto, Browne, and Bockelman, Phys. Rev. **91**, 1502 (1953).
Bu 54 W. W. Buechner, Proceedings of the Glasgow Conference, Pergamon Press, London (1954).
Bu 56 Buechner, Mazari, and Sperduto, Phys. Rev. **101**, 188 (1956).
Bu 56a Burge, Fujimoto, and Hossain, Phil. Mag. **1**, 19 (1956).
Bu 56b Bumiller, Müller, and Staub, Helv. Phys. Acta. **29**, 234(A) (1956), and **29**, 83 (1956).
Bu 56c B. S. Burton and R. M. Williamson, Bull. Am. Soc. Ser. II, **1**, 264 (1956).
Bu 56d Burrows, Green, Hinds, and Middleton, Proc. Phys. Soc. (London) **A69**, 310 (1956).

Bu 57 W. W. Buechner and A. Sperduto, Phys. Rev. **106**, 1008 (1957).
Ca 53 H. Casson, Phys. Rev. **89**, 809 (1953).
Ca 54 U. Cappeller and R. Klingelhöfer, Z. Naturforsch. **9A**, 1052(L) (1954).
Ca 54a Carlson, Geer, and Nelson, Phys. Rev. **94**, 1311 (1954).
Ca 55 Calvert, Yaffe, Litherland, and Maslin, Proc. Phys. Soc (London) **A68**, 1008 (1955).
Ca 56 P. J. Campion and G. A. Bartholomew, Bull. Am. Phys. Soc. Ser. II, **1**, 28 (1956) plus verbal report to Nuclear Data Group.
Ca 56a Calvert, Yaffe, and Maslin, Phys. Rev. **101**, 501(L) (1956).
Ca 57 Calvert, Yaffe, and Maslin, Proc. Phys. Soc. (London) **A70**, 78 (1957).
Ch 53 Churchill, Jones, and Hunt, Nature **172**, 460(L) (1953).
Ch 53a L. S. Cheng and M. L. Pool, Phys. Rev. **90**, 886 (1953).
Ch 54 G. Charpak and F. Suzor, J. phys. radium **15**, 378 (1954).
Ch 55 G. Charpak, J. phys. radium **16**, 62 (1955).
Ch 55a G. Charpak, J. phys. radium **16**, 567 (1955).
Ch 56 Chick, Evans, Hancock, Hunt, and Pope, Proc. Phys. Soc. (London) **A69**, 624 (1956).
Cl 55 Clarke, Almqvist, and Paul, Phys. Rev. **99**, 654(A) (1955).
Co 40 J. M. Cork and W. Middleton, Phys. Rev. **58**, 474(L) (1940).
Co 50 F. C. W. Colmer and D. J. Littler, Proc. Phys. Soc. (London) **A63**, 1175 (1950).
Co 50a J. W. Cobble and R. W. Atteberry, Phys. Rev. **80**, 917(L) (1950).
Co 53 Cork, LeBlanc, Brice, and Nester, Phys. Rev. **92**, 367 (1953).
Co 54 B. L. Cohen, Phys. Rev. **94**, 117 (1954).
Co 54a Cohen, Reynolds, and Zucker, Phys. Rev. **96**, 1617 (1954).
Co 54b Cox, Van Loef, and Lind, Phys. Rev. **93**, 925(A) (1954).
Co 55 H. E. Conzett, Phys. Rev. **100**, 1794(A) (1955).
Co 55a Cohn, Bair, Kington, and Willard, Phys. Rev. **99**, 644(A) (1955) plus verbal report to Nuclear Data Group.
Co 55b B. L. Cohen, Phys. Rev. **98**, 49 (1955).
Co 55c C. F. Coleman, Phil. Mag. **46**, 1135(L) (1955).
Co 56 A. V. Cohen and P. H. White, Nuclear Physics **1**, 73 (1956).
Co 56a S. A. Cox and R. M. Williamson, Bull. Am. Phys. Soc. Ser. II, **1**, 196 (1956).
Co 56b L. Colli and U. Facchini, Physica **22**, 1158(A) (1956).
Co 57 W. R. Cobb and D. B. Guthe, Phys. Rev. **107**, 181 (1957).
Cr 40 Creutz, Fox, and Sutton, Phys. Rev. **57**, 567(A) (1940).
Cr 55 W. G. Cross and R. G. Jarvis, Phys. Rev. **99**, 621(A) (1955).
Cr 56 D. S. Craig, Phys. Rev. **101**, 1479 (1956).
Cr 56a L. Cranberg and J. S. Levin, Phys. Rev. **103**, 343 (1956).
Cr 57 Crutchfield, Haeberli, and Newson, Bull. Am. Phys. Soc. Ser. II, **2**, 33 (1957).
Cu 40 Curran, Dee, and Strothers, Proc. Roy. Soc. (London) **A174**, 546 (1940).
Cu 56 Culler, Fernbach, and Sherman, Phys. Rev. **101**, 1047 (1956).
Da 49 Davison, Buchanan, and Pollard, Phys. Rev. **76**, 890 (1949).
Da 53 Daniel, Koester, and Mayer-Kuckuk, Z. Naturforsch. **8a**, 447(L) (1953).
Da 54 I. E. Dayton, Phys. Rev. **95**, 754 (1954).
Da 54a H. Daniel and W. Bothe, Z. Naturforsch. **9a**, 402 (1954).
Da 55 R. Davis, Phys. Rev. **97**, 766 (1955).
Da 56 Day, Johnsrud, and Lind, Bull. Am. Phys. Soc. Ser. II, **1**, 56 (1956) plus verbal report to Nuclear Data Group.
Da 56a I. E. Dayton and G. Schrank, Phys. Rev. **101**, 1358 (1956).
Da 56b R. Davis, Bull. Am. Phys. Soc. Ser. II, **1**, 219 (1956).
Da 56c R. B. Day, Phys. Rev. **102**, 767 (1956).
Da 56d W. K. Dawson, Can. J. Phys. **34**, 1480 (1956).
Da 57 Dalton, Hinds, and Parry, Proc. Phys. Soc. (London) **A70**, 586 (1957).
De 53 C. F. G. Delaney and J. H. J. Poole, Phys. Rev. **89**, 529(L) (1953).
De 53a A. De-Shalit, Phys. Rev. **91**, 1479 (1953).
De 53b A. De-Shalit and M. Goldhaber, Phys. Rev. **92**, 1211 (1953).
De 55 M. D. de Souza Santos *et al.*, Geneva Conference Report 8/P/897 (1955).
De 56 G. Dearnaley, Phil. Mag. **1**, 821 (1956).
De 56a Devons, Manning, and Towle, Proc. Phys. Soc. (London) **A69**, 173 (1956).
De 57 J. de Veiga Simao and J. P. F. Sellschop (to be published).
Di 55 Dixon, McNair, and Curran, J. phys. radium **16**, 538 (1955).
Do 53 Donahue, Jones, McEllistrem and Richards, Phys. Rev. **89**, 824 (1953).
Do 56 W. T. Doyle and A. B. Robbins, Phys. Rev. **101**, 1056 (1956).
Do 56a Dolan, Fincher, Kenny, Berko, and Whitehead, Bull. Am. Phys. Soc. Ser. II, **1**, 339 (1956).
Dr 55 R. W. P. Drever and A. Moljk, Phil. Mag. **46**, 1337 (1955).
Dr 56 B. J. Dropesky and A. W. Schardt, Phys. Rev. **102**, 426 (1956).
El 41 D. R. Elliot and L. D. P. King, Phys. Rev. **60**, 489 (1941).
El 52 L. G. Elliot, Phys. Rev. **85**, 942(L) (1952).
El 54 Elbek, Nielsen, and Nielsen, Phys. Rev. **95**, 96 (1954).
El 55 Elbek, Madsen, and Nathan, Phil. Mag. **46**, 663 (1955).
El 55a F. A. El Bedewi and M. A. El Wahab, Proc. Phys. Soc. (London) **A68**, 754 (1955).
El 56 J. O. Elliot, Phys. Rev. **101**, 684 (1956).
El 56a M. El-Wahab, thesis, Alexandria (1956).
Em 54 Emmerich, Singer, and Kurbatov, Phys. Rev. **94**, 113 (1954).
Em 55 E. W. Emery and N. Veall, Proc. Phys. Soc. (London) **A68**, 346(L) (1955).
En 51 Enge, Buechner, Sperduto, and Van Patter, Phys. Rev. **83**, 31 (1951).
En 51a Endt, Van Patter, Buechner, and Sperduto, Phys. Rev. **83**, 491 (1951).
En 51b W. W. Ennis, Phys. Rev. **82**, 204(A) (1951).
En 52 Endt, Haffner, and Van Patter, Phys. Rev. **86**, 518 (1952).
En 52a Enge, Buechner, and Sperduto, Phys. Rev. **88**, 963 (1952).
En 52b Endt, Enge, Haffner, and Buechner, Phys. Rev. **87**, 27 (1952).
En 53 T. C. Engelder, Phys. Rev. **90**, 259 (1953).
En 53a H. A. Enge, Universitetet i Bergen Årbok (1953).
En 54 A. A. Enge, Phys. Rev. **94**, 730(L) (1954).
En 54a P. M. Endt and J. C. Kluyver, Revs. Modern Phys. **26**, 95 (1954).
En 54b Endt, Kluyver, and van der Leun, Phys. Rev. **95**, 580(L) (1954).
En 54c Endt, Kluyver, and van der Leun, Physica **20**, 1299 (1954).
En 56 Enge, Angleman, and Jarrell, M.I.T., Laboratory for Nuclear Science Progress Report, May 31 (1956).
En 56a Enge, Buechner, Sperduto, and Mazari, Bull. Am. Phys. Soc. Ser. II, **1**, 212 (1956).
En 56b P. M. Endt, Physica **22**, 1062 (1956).
En 56c Endt, Paris, Sperduto, and Buechner, Phys. Rev. **103**, 961 (1956).
En 57 Endt, Buechner, Braams, Paris, and Sperduto, Phys. Rev. **105**, 1002 (1957).
En 57a P. M. Endt and C. H. Paris, Phys. Rev. **106**, 764 (1957).
Fa 50 W. R. Faust, Phys. Rev. **78**, 624(L) (1950).
Fa 55 Farney, Givin, Kern, and Hahn, Phys. Rev. **97**, 720 (1955).
Fe 52 L. Feldman and C. S. Wu, Phys. Rev. **87**, 1091 (1952).
Fe 53 A. J. Ferguson and H. E. Gove, Phys. Rev. **91**, 439(A) (1953).
Fe 54 Ferguson, Halpern, Nathans, and Yergin, Phys. Rev. **95**, 776 (1954).
Fe 54a L. Feuvrais and T. Yuasa, Compt. rend. **239**, 1627 (1954).
Fe 55 E. Feenberg, Phys. Rev. **99**, 71 (1955).

Fe 56 J. M. Ferguson, M.I.T., Laboratory for Nuclear Science Progress Report, May 31 (1956).
Fi 51 R. E. Fields and M. Walt, Phys. Rev. **83**, 479(L) (1951).
Fi 54 G. E. Fischer, Phys. Rev. **96**, 704 (1954).
Fl 54 Flack, Rutherglen, and Grant, Proc. Phys. Soc. (London) **A67**, 973 (1954).
Fo 52 A. Folkierski, Proc. Phys. Soc. (London) **A65**, 1006 (1952).
Fo 54 G. N. Fowler, Proc. Phys. Soc. (London) **A67**, 1005(L) (1954).
Fo 56 G. Ford and R. R. Lewis, Bull. Am. Phys. Soc. Ser. II, **1**, 195 (1956).
Fr 50 J. M. Freeman, Proc. Phys. Soc. (London) **A63**, 668(L) (1950).
Fr 50a A. P. French and P. B. Treacy, Proc. Phys. Soc (London) **A63**, 665 (1950).
Fr 51 J. M. Freeman and J. Seed, Proc. Phys. Soc. (London) **A64**, 313(L) (1951).
Fr 54 Freemantle, Prowse, Hossain, and Rotblat, Phys. Rev. **96**, 1270 (1954).
Fr 55 J. B. French and B. J. Raz, Phys. Rev. **98**, 1523(L) (1955).
Fu 48 H. W. Fulbright and R. R. Bush, Phys. Rev. **74**, 1323 (1948).
Fu 51 H. W. Fulbright and J. C. D. Milton, Phys. Rev. **82**, 274(L) (1951).
Fu 54 Fujimoto, Kikuchi, and Yoshida, Progr. Theoret. Phys. (Kyoto) **11**, 264 (1954).
Ga 53 Galonsky, Haeberli, Goldberg, and Douglas, Phys. Rev. **91**, 439(A) (1953).
Ge 55 Gerber, Garcia, Muñoz, and Maeder, Helv. Phys. Acta **28**, 478 (1955).
Ge 56 Gerber, Garcia, Muñoz, and Maeder, Phys. Rev. **101**, 774 (1956).
Ge 56a Gerstein, Niederer, and Strauch, Bull. Am. Phys. Soc. Ser. II, **1**, 192 (1956).
Ge 56b S. Geschwind (private communication).
Gi 44 Gibert, Roggen, and Rossel, Helv. Phys. Acta **17**, 97 (1944).
Gi 52 W. M. Gibson and E. E. Thomas, Proc. Roy. Soc. (London) **A210**, 543 (1952).
Gi 52a D. A. Gilbert, Phys. Rev. **85**, 716(A) (1952).
Gl 47 E. Gleditsch and F. Graf, Phys. Rev. **72**, 640(L) (1947).
Gl 55 N. W. Glass and J. R. Richardson, Phys. Rev. **98**, 1251 (1955).
Gl 56 R. J. Glauber and P. C. Martin, Phys. Rev. **104**, 158 (1956).
Go 50 H. E. Gove, M.I.T., Laboratory for Nuclear Science and Engineering Progress Report, July (1950).
Go 53 E. Goldberg, Phys. Rev. **89**, 760 (1953).
Go 53a H. E. Gove (private communication).
Go 54 Gove, Paul, Bartholomew, and Litherland, Phys. Rev. **94**, 749(A) (1954).
Go 54a J. Goldemberg and L. Katz, Can. J. Phys. **32**, 49 (1954).
Go 54b M. Goodrich and W. B. Payne, Phys. Rev. **94**, 405 (1954).
Go 54c Goldberg, Haeberli, Galonsky, and Douglas, Phys. Rev. **93**, 799 (1954).
Go 55 Gorodetzky, Muller, and Port, Compt. rend. **240**, 1704 and 2224 (1955).
Go 56 Gove, Bartholomew, Paul, and Litherland, Nuclear Physics **2**, 132 (1956).
Go 56a S. Goldstein and I. Talmi, Phys. Rev. **102**, 589(L) (1956).
Go 56b P. Goldhammer, Phys. Rev. **101**, 1375 (1956).
Go 57 Gove, Litherland, and Paul, Bull. Am. Phys. Soc. Ser. II, **2**, 178 (1957).
Go 57a Gove, Bartholomew, Paul, and Litherland, Nuclear Physics (erratum) (to be published).
Go 57b H. E. Gove (private communication).
Gr 50 Grotdal, Lönsjö, Tangen, and Bergström, Phys. Rev. **77**, 296(L) (1950).
Gr 50a P. J. Grant, Proc. Phys. Soc. (London) **A63**, 1298 (1950).
Gr 50b J. C. Grosskreutz and K. B. Mather, Phys. Rev. **77**, 580 (1950).
Gr 51 G. R. Grove and J. N. Cooper, Phys. Rev. **82**, 505 (1951).
Gr 54 Green, Harris, and Cooper, Phys. Rev. **96**, 817(A) (1954).
Gr 54a G. W. Greenlees, Proc. Phys. Soc. (London) **A67**, 1107(L) (1954).
Gr 55 G. W. Greenlees, Proc. Phys. Soc. (London) **A68**, 97 (1955).
Gr 55a Groshev, Adyasevich, and Demidov, International Conference Geneva Report 8/P/651 (1955).
Gr 55b E. R. Graves and R. W. Davis, Phys. Rev. **97**, 1205 (1955).
Gr 55c Green, Singh, and Willmott, Phil. Mag. **46**, 982 (1955).
Gr 55d G. L. Griffith, Phys. Rev. **98**, 579 (1955).
Gr 55e Grant, Rutherglen, Flack, and Hutchinson, Proc. Phys. Soc. (London) **A68**, 369 (1955).
Gr 55f T. S. Green, thesis, Liverpool (1955).
Gr 56 D. Green and J. R. Richardson, Phys. Rev. **101**, 776 (1956).
Gr 56a W. E. Graves and S. K. Suri, Phys. Rev. **101**, 1368 (1956).
Gr 56b J. S. Greenberg and M. Deutsch, Phys. Rev. **102**, 415 (1956).
Gr 56c W. C. Grayson and L. W. Nordheim, Phys. Rev. **102**, 1093 (1956).
Gr 56d H. G. Graetzer and A. B. Robbins, Bull. Am. Phys. Soc. Ser. II, **1**, 280 (1956).
Gr 56e Green, Singh, and Willmott, Proc. Phys. Soc. (London) **A69**, 335 (1956).
Gr 56f G. W. Greenlees and A. E. Souch, Proc. Phys. Soc. (London) **A69**, 686 (1956).
Gu 54 P. C. Gugelot, Phys. Rev. **93**, 425 (1954).
Gu 56 P. C. Gugelot and M. Rickey, Phys. Rev. **101**, 1613(L) (1956).
Gu 56a P. C. Gugelot and P. R. Phillips, Phys. Rev. **101**, 1614(L) (1956).
Ha 35 O. Haxel, Z. tech. Physik **16**, 400 (1935).
Ha 50 Harris, Muehlhause, Rasmussen, Schroeder, and Thomas, Phys. Rev. **80**, 342 (1950).
Ha 50a Haslam, Katz, Moody, and Skarsgard, Phys. Rev. **80**, 318 (1950).
Ha 51 T. D. Hanscome and C. W. Malich, Phys. Rev. **82**, 304(A) (1951).
Ha 51a W. Hälg, Helv. Phys. Acta **24**, 641(A) (1951).
Ha 52 Hausman, Allen, Arthur, Bender, and McDole, Phys. Rev. **88**, 1296 (1952).
Ha 52a Haslam, Summers-Gill, and Crosby, Can. J. Phys. **30**, 257 (1952).
Ha 52b B. Hamermesh and V. Hummel, Phys. Rev. **88**, 916 (1952).
Ha 52c J. A. Harvey, Phys. Rev. **88**, 162(A) (1952).
Ha 53 Haeberli, Galonsky, Goldberg, and Douglas, Phys. Rev. **91**, 438(A) (1953).
Ha 53a Hamilton, Lemonick, and Pipkin, Phys. Rev. **92**, 1191 (1953).
Ha 54 Halpern, Nathans, and Yergin, Phys. Rev. **95**, 1529 (1954).
Ha 54a Haslam, Roberts, and Robb, Can. J. Phys. **32**, 361 (1954).
Ha 54b J. Haffner, M.I.T., Laboratory for Nuclear Science Progress Report, August 31, 1954.
Ha 55 Hattori, Hisatake, Mikumo, and Momata, J. Phys. Soc. Japan **10**, 242 (1955).
Ha 55a W. Haeberli, Phys. Rev. **99**, 640(A) (1955).
Ha 55b T. H. Handley and W. S. Lyon, Phys. Rev. **99**, 755 (1955).
Ha 55c D. A. Hancock and F. Verdaguer, Proc. Phys. Soc. (London) **A68**, 1080 (1955).
Ha 56 Hausman, Monahan, Mooring, and Raboy, Bull. Am. Phys. Soc. Ser. II, **1**, 56 (1956) plus verbal report to Nuclear Data Group.
Ha 56a R. K. Haling, Bull. Am. Phys. Soc. Ser. II, **1**, 29 (1956).
Ha 56b Hahn, Revenhall, and Hofstadter, Phys. Rev. **101**, 1131 (1956).
Ha 56c J. W. Haffner, Phys. Rev. **103**, 1398 (1956).
Ha 56d H. E. Handler and J. R. Richardson, Phys. Rev. **102**, 833 (1956).
Ha 56e Hassler, Steigert, and Pieper, Bull. Am. Phys. Soc. Ser. II, **1**, 280 (1956).
He 47 Heitler, May, and Powell, Proc. Roy. Soc. (London) **A190**, 180 (1947).

He 49 Herb, Snowdon, and Sala, Phys. Rev. **75**, 246 (1949).
He 52 A. Hedgran and D. Lind, Arkiv Fysik **5**, 177 (1952).
He 54 J. Heintze, Z. Naturforsch. **9a**, 469 (1954).
He 54a Hee, Coche, Keller, Jarovoy, and Wack, Ann. geophys. **10**, 19 (1954).
He 54b N. P. Heydenberg and G. M. Temmer, Phys. Rev. **94**, 1252 (1954).
He 56 R. H. Helen, Phys. Rev. **104**, 1466 (1956).
Hi 45 Hibdon, Pool, and Kurbatov, Phys. Rev. **67**, 289 (1945).
Hi 54 A. Hitchcock, Phil. Mag. **45**, 379 and 385 (1954).
Hi 54a Hird, Whitehead, Butler, and Collie, Phys. Rev. **96**, 702 (1954).
Hi 55 N. M. Hintz, Phys. Rev. **100**, 1794(A) (1955).
Hi 56 N. M. Hintz and R. M. Eisberg, Phys. Rev. **103**, 645 (1956).
Hi 56a R. W. Hill, Bull. Am. Phys. Soc. **1**, 174 (1956).
Hi 56b S. Hinds and R. Middleton, Proc. Phys. Soc. (London) **A69**, 347(L) (1956).
Ho 40 J. B. Hoag, Phys. Rev. **57**, 937(L), 1940.
Ho 46 N. Hole and K. Siegbahn, Arkiv Mat. Astron. Fysik **33A**, No. 9 (1946).
Ho 50 Houtermans, Haxel, and Heintze, Z. Physik **128**, 657 (1950).
Ho 53 J. R. Holt and T. N. Marshan, Proc. Phys. Soc. (London) **A66**, 258 (1953).
Ho 53a J. R. Holt and T. N. Marsham, Phys. Rev. **89**, 665(L) (1953).
Ho 53b N. F. Hornyak and T. Coor, Phys. Rev. **92**, 675 (1953).
Ho 53c J. R. Holt and T. N. Marsham, Proc. Phys. Soc. (London) **A66**, 565 (1953).
Ho 53d J. R. Holt and T. N. Marsham, Proc. Phys. Soc. (London) **A66**, 467 (1953).
Ho 56 Hoyle, Fowler, Burbidge, and Burbidge, Science **124**, 611 (1956).
Ho 57 Hoogenboom, Van Loef, and Endt (unpublished).
Hu 41 P. Huber, Helv. Phys. Acta **14**, 163 (1941).
Hu 45 Huber, Lienhard, Scherrer, and Wäffler, Helv. Phys. Acta **18**, 221 (1945).
Hu 53 S. E. Hunt and W. M. Jones, Phys. Rev. **89**, 1283 (1953).
Hu 54 Hunt, Kline, and Zaffarano, Phys. Rev. **95**, 611(A) (1954).
Hu 54a W. A. Hunt and D. J. Zaffarano, U.S.A.E.C. Report ISC—469 (1954).
Hu 54b Hunt, Jones, Churchill, and Hancock, Proc. Phys. Soc. (London) **A67**, 443 (1954).
Hu 54c Hunt, Jones, and Churchill, Proc. Phys. Soc. (London) **A67**, 479(L) (1954).
Hu 55 S. E. Hunt and D. A. Hancock, Phys. Rev. **97**, 567(L) (1955).
Hu 55a T. Hürlimann and P. Huber, Helv. Phys. Acta **28**, 33 (1955).
Hu 55b R. Huby and H. C. Newns, Appendix to El 55a.
Hu 55c D. J. Hughes and J. A. Harvey, Brookhaven National Laboratory Report BNL 325 (1955).
Hu 56 I. S. Hughes and D. Sinclair, Proc. Phys. Soc. (London) **A69**, 125 (1956).
Hu 56a K. Huang, Phys. Rev. **102**, 422 (1956).
Ig 56 Igo, Wegner, and Eisberg, Phys. Rev. **101**, 1508 (1956).
In 50 Inghram, Brown, Patterson, and Hess, Phys. Rev. **80**, 916(L) (1950).
In 53 D. R. Inglis, Revs. Modern Phys. **25**, 390 (1953).
It 41 J. Itoh, Proc. Phys. Math. Soc. Japan **23**, 605 (1941).
Iw 53 Iwersen, Koski, and Rasetti, Phys. Rev. **91**, 1229 (1953).
Iw 55 J. E. Iwersen and W. S. Koski, Phys. Rev. **98**, 1307 (1955).
Je 52 Jensen, Nichols, Clement, and Pohm, Phys. Rev. **85**, 112 (1952).
Je 53 C. D. Jeffries, Phys. Rev. **90**, 1130(L) (1953).
Jo 49 F. Johnston and J. E. Willard, Phys. Rev. **75**, 528(L) (1949).
Jo 51 Johnson, Gordy, and Livingston, Phys. Rev. **83**, 1249(L) (1951).
Jo 52 J. W. Jones and T. P. Kohman, Phys. Rev. **85**, 941(L) (1952).
Jo 52a W. H. Johnson, Phys. Rev. **88**, 1213(L) (1952).
Jo 53 J. W. Jones and T. R. Kohman, Phys. Rev. **90**, 495(L) (1953).
Jo 54 G. A. Jones and D. H. Wilkinson, Phil. Mag. **45**, 230(L) (1954).
Jo 55 S. A. E. Johansson, Phys. Rev. **97**, 1186(L) (1955).
Jo 55a Joki, Miller, and Evans, Phys. Rev. **99**, 610(A) (1955).
Jo 56 Johnson, Johnson, and Langer, Phys. Rev. **102**, 1142 (1956).
Ju 56 A. C. Juveland and W. Jentschke, Bull. Am. Phys. Soc. Ser. II, **1**, 193 (1956).
Ka 51 L. Katz and A. G. W. Cameron, Phys. Rev. **84**, 1115 (1951).
Ka 51a L. Katz and A. S. Penfold, Phys. Rev. **81**, 815 (1951).
Ka 52 Kaufmann, Goldberg, Koester, and Mooring, Phys. Rev. **88**, 673 (1952).
Ka 52a S. Katcoff, Phys. Rev. **87**, 886 (1952).
Ka 54 Katz, Haslam, Goldemberg, and Taylor, Can. J. Phys. **32**, 580 (1954).
Ka 55 Kavanagh, Mills, and Sherr, Phys. Rev. **97**, 248(L) (1955).
Ke 54 Kelley, Kuhn, and Pery, Proc. Phys. Soc. (London) **A67**, 450 (1954).
Ke 56 O'Kelley, Lazar, and Eichler, Phys. Rev. **101**, 1059 (1956).
Ke 56a B. D. Kern and L. W. Cochran, Phys. Rev. **104**, 711 (1956).
Ki 42 Kikuchi, Watase, and Itoh, Z. Physik **69**, 185 (1942).
Ki 51 Kinsey, Bartholomew, and Walker, Phys. Rev. **83**, 519 (1951).
Ki 52 Kinsey, Bartholomew, and Walker, Phys. Rev. **85**, 1012 (1952).
Ki 52a J. S. King and W. C. Parkinson, Phys. Rev. **88**, 141(L) (1952).
Ki 52b B. B. Kinsey and G. A. Bartholomew, Physica **18**, 1112 (1952).
Ki 53 Kington, Bair, Carlson, and Willard, Phys. Rev. **89**, 530(L) (1953).
Ki 53a B. B. Kinsey and G. A. Bartholomew, Can. J. Phys. **31**, 901(L) (1953).
Ki 53b J. S. King and E. H. Beach, Phys. Rev. **90**, 381(A) (1953).
Ki 53c B. B. Kinsey and G. A. Bartholomew, Can. J. Phys. **31**, 537 (1953).
Ki 54 R. M. Kiehn and C. Goodman, Phys. Rev. **95**, 989 (1954).
Ki 54a B. B. Kinsey and G. A. Bartholomew, Phys. Rev. **93**, 1260 (1954).
Ki 54b R. W. King, Revs. Modern Phys. **26**, 327 (1954).
Ki 55 R. W. King, Phys. Rev. **99**, 67 (1955).
Ki 55a Kington, Bair, Cohn, and Willard, Phys. Rev. **99**, 1393 (1955).
Ki 56 Kistner, Schwarzschild, and Rustad, Phys. Rev. **104**, 154 (1956).
Kl 54 Kluyver, van der Leun, and Endt, Physica **20**, 1287 (1954).
Kl 54a Kluyver, van der Leun, and Endt, Phys. Rev. **94**, 1795(L) (1954).
Kl 54b R. M. Kline and D. J. Zaffarano, Phys. Rev. **96**, 1620 (1954).
Kl 55 J. C. Kluyver and C. van der Leun, Physica **21**, 604(L) (1955).
Kl 55a J. C. Kluyver and P. M. Endt, Conference on Electromagnetically Enriched Isotopes Harwell (1955).
Kl 57 Kloepper, Day, and Lind, Bull. Am. Phys. Soc. Ser. II, **2**, 60 (1957).
Ko 52 L. J. Koester, Phys. Rev. **85**, 643 (1952).
Ko 54 Koerts, Schwarzschild, Gold, and Wu, Phys. Rev. **95**, 612(A) (1954).
Ko 54a L. Koester, Z. Naturforsch. **9a**, 104 (1954).
Ko 55 S. Kono, J. Phys. Soc. Japan **10**, 495 (1955).
Kr 53 A. Z. Kranz and W. W. Watson, Phys. Rev. **91**, 1472 (1953).
Kr 53a L. M. Khromchenko, Doklady Akad. Nauk SSSR **93**, 451 (1953).
Kr 54 W. E. Kreger, Phys. Rev. **96**, 1554 (1954).
Kr 54a Kraushaar, Mihelich, and Sunyar, Phys. Rev. **95**, 456 (1954).
Kr 54b L. M. Khromchenko, Doklady Akad. Nauk SSSR **94**, 1037 (1954).

Kr 56 R. W. Krone, A. Everett, and S. S. Hanna, Bull. Am. Phys. Soc. Ser. II, 1, 329 (1956).
Kr 57 Kruse, Bent, and Eklund, Bull. Am. Phys. Soc. Ser. II, 2, 29 (1957).
Ku 53 D. Kurath, Phys. Rev. 91, 1430 (1953).
La 39 Langer, Mitchell, and McDaniel, Phys. Rev. 56, 962(L) (1939).
La 39a J. L. Lawson, Phys. Rev. 56, 131 (1939).
La 48 R. V. Langmuir, Phys. Rev. 74, 1559(A) (1948).
La 49 L. M. Langer and H. C. Price, Phys. Rev. 76, 641 (1949).
La 50 L. M. Langer, Phys. Rev. 77, 50 (1950).
La 51 H. H. Landon, Phys. Rev. 83, 1081 (1951).
La 54 Langevin, Yuasa, and Mérinis, J. phys. radium 15, 778 (1954).
La 54a H. Langevin and N. Marty, J. phys. radium 15, 127 (1954).
La 55 M. J. Laubitz, Proc. Phys. Soc. (London) A68, 1033 (1955).
La 55a M. Langevin and P. Radvanyi, Compt. rend. 241, 33 (1955).
La 56 R. O. Lane and J. E. Monahan, Bull. Am. Phys. Soc. Ser. II, 1, 187 (1956).
La 56a R. O. Lane and J. E. Monahan, Bull. Am. Phys. Soc. Ser. II, 1, 346 (1956).
La 56b Laberrigue-Frolow, Radvanyi, and Langevin, J. phys. radium 17, 530 (1956).
La 57 Lazar, Lyon, and Macklin, Bull. Am. Phys. Soc. Ser. II, 2, 15 (1957).
Le 54 R. D. Leamer and G. W. Hinman, Phys. Rev. 96, 1607 (1954).
Le 55 C. Levinson and K. W. Ford, Phys. Rev. 100, 13 (1955).
Le 56 Lewis, Olness, and Haeberli, Bull. Am. Phys. Soc. Ser. II, 1, 213 (1956).
Le 56a H. W. Lewis and W. T. Joyner, Bull. Am. Phys. Soc. Ser. II, 1, 280 (1956).
Le 56b L. L. Lee and F. P. Morring, Phys. Rev. 104, 1342 (1956).
Li 37 M. S. Livingston and H. A. Bethe, Revs. Modern Phys. 9, 245 (1937).
Li 53 M. Lindner, Phys. Rev. 89, 1150(L) (1953).
Li 53a M. Lindner, Phys. Rev. 91, 642 (1953).
Li 54 T. Lindqvist and A. C. G. Mitchell, Phys. Rev. 95, 1535 (1954).
Li 54a T. Lindqvist and A. C. G. Mitchell, Phys. Rev. 95, 444 (1954).
Li 55 K. Lidén and N. Starfelt, Phys. Rev. 97, 419 (1955).
Li 55a Little, Leonard, Prud'homme, and Vincent, Phys. Rev. 98, 634 (1955).
Li 55b T. Lindqvist and C. S. Wu, Phys. Rev. 100, 145 (1955).
Li 56 L. J. Lidofsky and V. K. Fischer, Phys. Rev. 104, 759 (1956).
Li 56a Litherland, Paul, Bartholomew, and Gove, Phys. Rev. 102, 208 (1956).
Li 57 A. E. Litherland *et al.* (private communication).
Ly 37 E. M. Lyman, Phys. Rev. 51, 1 (1937).
Ly 54 W. S. Lyon and J. J. Manning, Phys. Rev. 93, 501 (1954).
Ly 55 W. S. Lyon and T. H. Handley, Phys. Rev. 100, 1280 (1955).
Ly 56 W. S. Lyon and N. H. Lazar, Phys. Rev. 101, 1524 (1956).
Ma 37 C. Magnan, Compt. rend. 205, 1147 (1937).
Ma 41 C. Magnan, Ann. phys. 15, 5 (1941).
Ma 50 J. E. Mack, Revs. Modern Phys. 22, 64 (1950).
Ma 50a Macklin, Lidofsky, and Wu, Phys. Rev. 78, 318(A) (1950).
Ma 52 Mandeville, Swann, Chatterjee, and Van Patter, Phys. Rev. 85, 193 (1952).
Ma 52a W. H. Martin and S. W. Breckon, Can. J. Phys. 30, 643 (1952).
Ma 53 L. Marquez, Phys. Rev. 92, 1511 (1953).
Ma 53a L. Marquez, Phys. Rev. 90, 330(L) (1953).
Ma 54 Marty, Nataf, and Prentki, J. phys. radium 15, 134 (1954).
Ma 54a T. Mayer-Kuckuk, Z. Naturforsch. 9a, 338 (1954).
Ma 54b J. Mattauch and R. Bieri, Z. Naturforsch. 9, 303 (1954).
Ma 54c Maeder, Müller, and Wintersteiger, Helv. Phys. Acta 27, 3 (1954).
Ma 55 D. Maeder and P. Stähelin, Helv. Phys. Acta 28, 193 (1955).
Ma 55a Mazari, Alba, and Serment, Phys. Rev. 100, 972(A) (1955).
Ma 55b P. Macq, Bull. Acad. Roy. Méd. Belg. 41, 467 (1955).
Ma 56 Marion, Brugger, and Chapman, Phys. Rev. 101, 247 (1956).
Ma 56a Martin, Cork, and Burson, Phys. Rev. 102, 457 (1956).
Ma 56b Marion, Slattery, and Chapman, Phys. Rev. 103, 676 (1956).
Mc 35 E. McMillan and E. O. Lawrence, Phys. Rev. 47, 343 (1935).
Mc 40 McCreary, Kuerti, and Van Voorhis, Phys. Rev. 57, 351(A) (1940).
Mc 49 McElhinney, Hanson, Becker, Duffield, and Diven, Phys. Rev. 75, 542 (1949).
Mc 54 G. W. McClure, Phys. Rev. 94, 1637 (1954).
Mc 54a McPherson, Pederson, and Katz, Can. J. Phys. 32, 593 (1954).
Mc 55 J. A. McCarthy, Phys. Rev. 97, 1234 (1955).
Mc 56 McNair, Glover, and Wilson, Phil. Mag. 1, 199 (1956).
Me 37 A. Meye, Z. Physik. 105, 232 (1937).
Me 48 Metzger, Alder, and Huber, Helv. Phys. Acta 21, 278 (1948).
Me 54 W. E. Meyerhof and G. Lindstrom, Phys. Rev. 93, 949 (1954).
Me 56 Meadows, Diamond, and Sharp, Phys. Rev. 102, 190 (1956).
Me 57 Merritt, Campion, and Hawkings, Can. J. Phys. 35, 16 (1957).
Mi 50 Millar, Cameron, and Glicksmann, Can. J. Research 28A, 475 (1950).
Mi 52 C. Mileikowsky and W. Whaling, Phys. Rev. 88, 1254 (1952).
Mi 53 Middleton, El-Bedewi, and Tai, Proc. Phys. Soc. (London) A66, 95 (1953).
Mi 54 J. A. Miskel and M. L. Perlman, Phys. Rev. 94, 1683 (1954).
Mi 54a A. Michalowicz, J. phys. radium 15, 156 (1954).
Mi 55 Milani, Cooper, and Harris, Phys. Rev. 99, 645(A) (1955), plus verbal report to Nuclear Data Group.
Mi 56 A. B. Milojević, Bull. Inst. Nuclear Sci. Boris Kidrich 6, 21 (1956).
Mo 49 K. H. Morganstern and K. P. W. Wolf, Phys. Rev. 76, 1261 (1949).
Mo 50 H. T. Motz and R. F. Humphreys, Phys. Rev. 80, 595 (1950).
Mo 51 Mooring, Koester, Goldberg, Saxon, and Kaufmann, Phys. Rev. 84, 703 (1951).
Mo 51a P. Morrison, Phys. Rev. 82, 209 (1951).
Mo 52 H. T. Motz and D. E. Alburger, Phys. Rev. 86, 165 (1952).
Mo 52a H. T. Motz, Phys. Rev. 85, 501(L) (1952).
Mo 54 S. A. Moszkowski and D. C. Peaslee, Phys. Rev. 93, 455 (1954).
Mo 54a A. Moljk and S. C. Curran, Phys. Rev. 96, 395 (1954).
Mo 55 H. Morinaga, Phys. Rev. 97, 444 (1955).
Mo 55a H. Morinaga, Phys. Rev. 97, 1185(L) (1955).
Mo 55b I. L. Morgan, Phys. Rev. 99, 621(A) (1955).
Mo 55c H. Morinaga, Phys. Rev. 100, 431(L) (1955).
Mo 56 H. Morinaga and E. Bleuler, Bull. Am. Phys. Soc. Ser. II, 1, 30 (1956) plus verbal report to Nuclear Data Group.
Mo 56a H. T. Motz, Phys. Rev. 104, 1353 (1956).
Mo 56b H. Morinaga, Phys. Rev. 103, 504(L) (1956).
Mo 56c I. L. Morgan, Phys. Rev. 103, 1031 (1956) plus verbal report to Nuclear Data Group.
Mo 56d H. Morinaga and E. Bleuler, Phys. Rev. 103, 1423 (1956).
Mu 56 Th. Muller, thesis, Strassbourg, Ann. Phys. (to be published).
My 49 V. Myers and A. Wattenberg, Phys. Rev. 75, 992(L) (1949).
Na 53 M. Nahmias and T. Yuasa, Compt. rend. 236, 2399 (1953).
Na 53a J. Nagy, Acta Physica Acad. Sci. Hung. 3, 15 (1953).
Na 54 R. Nathans and J. Halpern, Phys. Rev. 93, 437 (1954).
Na 54a M. E. Nahmias, J. phys. radium 15, 568 (1954).
Na 55 R. Nathans and P. F. Yergin, Phys. Rev. 98, 1296 (1955).

Na 56 M. E. Nahmias and T. Yuasa, J. phys. radium 17, 373(L) (1956).
Na 56a M. E. Nahmias and T. Yuasa, J. phys. radium 17, 536 (1956).
Na 57 Nakada, Anderson, Gardner, and Wong, Bull. Am. Phys. Soc. Ser. II, 2, 32 (1957).
Ne 54 J. O. Newton, Phys. Rev. 96, 241(L) (1954).
Ne 56 Nemilov, Zherebtsova, and Funstein, Physica 22, 1155(A) (1956).
Ne 57 J. O. Newton (private communication).
Ni 50 A. O. Nier, Phys. Rev. 77, 789 (1950).
Ni 54 R. T. Nichols and E. N. Jensen, Phys. Rev. 94, 369 (1954).
Ni 55 R. T. Nichols and E. N. Jensen, Phys. Rev. 100, 1407 (1955).
Ny 55 K. Nybö and T. Grotdal, Nature 175, 130(L) (1955).
Og 54 R. A. Ogg Jr. and J. D. Ray, J. Chem. Phys. 22, 147 (1954).
Ol 54 J. L. Olsen and G. D. O'Kelley, Phys. Rev. 93, 1125(L) (1954).
Ol 55 J. W. Olness and H. W. Lewis, Phys. Rev. 99, 654(A) (1955).
Ol 56 Olness, Haeberli, and Lewis, Bull. Am. Phys. Soc. Ser. II, 1, 212 (1956) plus verbal report to Nuclear Data Group.
Ov 49 Overstreet, Jacobson, and Stout, Phys. Rev. 75, 231 (1949).
Ov 56 O. E. Overseth, Bull. Am. Phys. Soc. Ser. II, 1, 281 (1956).
Pa 52 Parkinson, Beach, and King, Phys. Rev. 87, 387(L) (1952).
Pa 52a M. Paganelli and G. Quareni, Phys. Rev. 86, 423(L) (1952).
Pa 54 Paul, Gove, Bartholomew, and Litherland, Phys. Rev. 749(A) (1954).
Pa 55 M. V. Pasechnik, International Conference Geneva Report 8/P/714 (1955).
Pa 55a Paul, Gove, Litherland, and Bartholomew, Phys. Rev. 1339 (1955).
Pa 55b Patterson, Newson, and Merzbacher, Phys. Rev. 99, 1625(A) (1955).
Pa 55c Paris, Buechner, and Endt, Phys. Rev. 100, 1317 (1955).
Pa 56 Paul, Bartholomew, Gove, and Litherland, Bull. Am. Phys. Soc. Ser. II, 1, 39 (1956).
Pa 56a S. P. Pandya, Phys. Rev. 103, 956 (1956).
Pa 57 Paris, van der Leun, and Endt, Bull. Am. Phys. Soc. Ser. II, 2, 179 (1957), and unpublished work.
Pa 57a S. P. Pandya and J. B. French, Bull. Am. Phys. Soc. Ser. II, 2, 27 (1957).
Pe 48 R. A. Peck, Phys. Rev. 73, 947 (1948).
Pe 49 R. A. Peck, Phys. Rev. 76, 1279 (1949).
Pe 50 Peterson, Barschall, and Bockelman, Phys. Rev. 79, 593 (1950).
Pe 53 D. C. Peaslee, Nuovo cimento 10, 1349(L) (1953).
Pe 56 Penning, Maltrud, Hopkins, and Schmidt, Phys. Rev. 104, 740 (1956).
Pe 57 J. R. Penning and F. H. Schmidt, Phys. Rev. 105, 647 (1957).
Pi 55 Pieper, Stanford, and Von Herrmann, Phys. Rev. 98, 1185(A) (1955).
Pl 40 Plain, Herb, Hudson, and Warren, Phys. Rev. 57, 187 (1940).
Pl 55 H. S. Plendl and F. E. Steigert, Phys. Rev. 98, 1538(A) (1955) plus verbal report to Nuclear Data Group.
Po 40 E. Pollard and W. W. Watson, Phys. Rev. 58, 12 (1940).
Po 49 Pollard, Sailor, and Wyly, Phys. Rev. 75, 725 (1949).
Po 49a Pontecorvo, Kirkwood, and Hanna, Phys. Rev. 75, 982(L) (1949).
Po 51 H. Pomerance, Phys. Rev. 83, 641 (1951).
Po 52 H. Pomerance, Phys. Rev. 88, 412 (1952).
Po 54 D. Popović, Nuovo cimento 12, 143(L) (1954).
Po 56 Pohm, Waddell, and Jensen, Phys. Rev. 101, 1315 (1956).
Pr 56 Frosser, Baumann, Brice, Read, and Krone, Phys. Rev. 104, 369 (1956).
Qu 56 A. R. Quinton and W. T. Doyle, Phys. Rev. 101, 669 (1956).
Qu 56a Quisenberry, Scolman, and Nier, Phys. Rev. 102, 1071 (1956).
Ra 53 N. F. Ramsey, *Nuclear Moments* (John Wiley and Sons, Inc., New York, 1953).
Ra 55 Rayburn, Lafferty, and Hahn, Phys. Rev. 98, 701 (1955).
Ra 55a Ramanna, Veeraraghaven, and Iyengar, Nuovo cimento 1, 623 (1955).
Ra 56 B. J. Raz and J. B. French, Bull. Am. Phys. Soc. Ser. II, 1, 223 (1956).
Ra 56a B. J. Raz, Bull. Am. Phys. Soc. Ser. II, 1, 336 (1956).
Re 52 Reilley, Allen, Arthur, Bender, Ely, and Hausman, Phys. Rev. 86, 857 (1952).
Re 52a C. Reutersward, Arkiv Fysik 4, 203 (1952).
Re 53 Reardon, Krone, and Stump, Phys. Rev. 91, 334 (1953).
Re 54 A. L. Recksiedler and B. Hamermesh, Phys. Rev. 96, 109 (1954).
Re 55 G. A. Renard, J. phys. radium 16, 575 (1955).
Re 56 J. B. Reynolds and K. G. Standing, Phys. Rev. 101, 158 (1956).
Re 56a C. Reutersward, Arkiv Fysik 11, 1 (1956).
Re 56b W. G. Read and R. W. Krone, Phys. Rev. 104, 1018 (1956).
Ri 48 H. T. Richards and R. V. Smith, Phys. Rev. 74, 1870(L) (1948).
Ri 50 Richards, Smith, and Browne, Phys. Rev. 80, 524 (1950).
Ri 55 L. S. Ring, Phys. Rev. 99, 137 (1955).
Ro 55 Roderick, Lönsjö, and Meyerhof, Phys. Rev. 97, 97 (1955).
Ro 55a Rothman, Hans, and Mandeville, Phys. Rev. 100, 83 (1955).
Ru 51 L. Ruby and J. R. Richardson, Phys. Rev. 83, 698 (1951).
Ru 52 Rudstam, Stevenson, and Folger, Phys. Rev. 87, 358 (1952).
Ru 52a Russel, Taylor, and Cooper, Phys. Rev. 86, 653(A) (1952).
Ru 53 J. G. Rutherglen and R. D. Smith, Proc. Phys. Soc. (London) A66, 800 (1953).
Ru 54 Rutherglen, Grant, Flack, and Deuchars, Proc. Phys. Soc. (London) A67, 101 (1954).
Ru 54a Russell, Taylor, and Cooper, Phys. Rev. 95, 99 (1954).
Ru 55 Rubin, Ajzenberg, and Reynolds, Phys. Rev. 98, 1185(A) (1955).
Ru 55a A. Russek, Phys. Rev. 99, 834 (1955).
Ru 55b R. A. Rubenstein and J. N. Snyder, Phys. Rev. 99, 189 (1955).
Ru 56 Rubin, Ajzenberg-Selove, and Mark, Phys. Rev. 104, 727 (1956).
Sa 49 V. L. Sailor, Phys. Rev. 75, 1836 (1949).
Sa 50 V. L. Sailor, Phys. Rev. 77, 794 (1950).
Sa 50a G. A. Sawyer and M. L. Wiedenbeck, Phys. Rev. 79, 490 (1950).
Sa 56 B. Saraf, Phys. Rev. 102, 466 (1956).
Sa 56a J. Sawicki, Physica 22, 1180(A) (1956).
Sc 48 Schelberg, Sampson, and Mitchell, Rev. Sci. Instr. 19, 458 (1948).
Sc 50 Schelberg, Sampson, and Cochran, Phys. Rev. 80, 574 (1950).
Sc 52 G. E. Schrank and J. R. Richardson, Phys. Rev. 86, 248(L) (1952).
Sc 52a Schoenfeld, Duborg, Preston, and Goodman, Phys. Rev. 85, 873 (1952).
Sc 54 G. Scharff-Goldhaber and M. McKeown, Phys. Rev. 95, 613(A) (1954).
Sc 54a C. Schwartz and A. De-Shalit, Phys. Rev. 94, 1257 (1954).
Sc 54b Schneider, Martin, Sempert, and Sutter, Helv. Phys. Acta 27, 172(A) (1954).
Sc 54c C. Schwartz, Phys. Rev. 94, 95 (1954).
Sc 54d Scherrer, Allison, and Faust, Phys. Rev. 96, 386 (1954).
Sc 55 J. P. Schiffer, Phys. Rev. 97, 428 (1954).
Sc 55a C. Schuhl and R. Basile, Compt. rend. 240, 2512 (1955).
Sc 56 Schwarzschild, Rustad, and Wu, Phys. Rev. 103, 1796 (1956).
Sc 56a M. J. Scott and R. E. Segel, Phys. Rev. 102, 1557 (1956).
Sc 56b Schwartz, Corbett, and Watson, Phys. Rev. 101, 1370 (1956).

Sc 56c Schiffer, Gossett, Phillips, and Young, Phys. Rev. **103**, 134 (1956).
Se 49 Seidlitz, Bleuler, and Tendam, Phys. Rev. **76**, 861(L) (1949).
Se 53 J. Seed, Phil. Mag. 44, 921(L) (1953).
Se 54 R. Sehr, Z. Physik **137**, 523 (1954).
Se 55 Seiler, Cooper, and Harris, Phys. Rev. **99**, 340(A) (1955).
Se 56 W. Selove, Phys. Rev. **101**, 231 (1956).
Se 56a J. C. Severiens and S. S. Hanna, Phys. Rev. **104**, 1612 (1956).
Sh 49 F. B. Shull and E. Feenberg, Phys. Rev. **75**, 1768(L) (1949).
Sh 51 Shoemaker, Faulkner, Bouricius, Kaufmann, and Mooring, Phys. Rev. **83**, 1011 (1951).
Sh 51a Sher, Halpern, and Mann, Phys. Rev. **84**, 387 (1951).
Sh 51b R. K. Sheline, Phys. Rev. **82**, 954 (1951).
Sh 51c Sheline, Holtzmark, and Fan, Phys. Rev. **83**, 919 (1951).
Sh 54 P. Shapiro, Phys. Rev. **93**, 290 (1954).
Sh 54a H. A. Shillibeer and R. D. Russel, Can. J. Phys. **32**, 681 (1954).
Sh 54b R. Sherr and R. H. Miller, Phys. Rev. **93**, 1076 (1954).
Sh 54c Sherr, Li, and Christy, Phys. Rev. **96**, 1258 (1954).
Sh 54d Sheline, Johnson, Bell, Davis, and McGowan, Phys. Rev. **94**, 1642 (1954).
Sh 56 Shaw, Conzett, Slobodrian, and Summers-Gill, Bull. Am. Phys. Soc. Ser. II, **1**, 253 (1956).
Sh 56a R. K. Sheline, Nuclear Physics **2**, 382 (1956).
Sh 57 H. A. Shillibeer, Geochim. Cosmochim. Acta (to be published).
Si 46 K. Siegbahn, Phys. Rev. **70**, 127 (1946).
Si 47 K. Siegbahn, Arkiv Mat. Astron. Fysik **34B**, No. 4 (1947).
Si 47a K. Siegbahn and A. Johansson, Arkiv. Mat. Astron. Fysik **34A**, No. 10 (1947).
Si 54 Simanton, Rightmire, Long, and Kohman, Phys. Rev. **96**, 1711(L) (1954).
Si 56 R. M. Sinclair, Westinghouse Research Report 60-94511-6-R 12 (1956).
Sm 42 G. P. Smith, Phys. Rev. **61**, 578 (1942).
Sm 51 R. D. Smith and R. A. Anderson, Nature **168**, 429(L) (1951).
Sm 54 Smith, Cooper, and Harris, Phys. Rev. **94**, 749(A) (1954) plus verbal report to Nuclear Data Group.
Sm 56 R. V. Smith, Bull. Am. Phys. Soc. Ser. II, **1**, 175 (1956).
Sm 57 R. T. Smith and D. Breitenbecher, Bull. Am. Phys. Soc. Ser. II, **2**, 59 (1957).
Sn 55 A. H. Snell and F. Pleasonton, Phys. Rev. **100**, 1396 (1955).
Sp 52 A. Sperduto and W. W. Buechner, Phys. Rev. **88**, 574 (1952).
Sp 55 A. Sperduto and W. W. Buechner, Phys. Rev. **100**, 961(A) (1955).
Sp 55a B. M. Spicer, Phys. Rev. **100**, 791 (1955).
Sp 55b A. Sperduto and W. W. Buechner, M.I.T., Laboratory for Nuclear Science Progress Report, Nov. 30 (1955).
St 50 R. M. Steffen, Phys. Rev. **80**, 115(L) (1950).
St 51 Strait, Van Patter, Buechner, and Sperduto, Phys. Rev. **81**, 747 (1951).
St 51a D. T. Stevenson and M. Deutsch, Phys. Rev. **83**, 1202 (1951).
St 51b D. T. Stevenson and M. Deutsch, Phys. Rev. **84**, 1071(L) (1951).
St 52 H. F. Stoddard and H. E. Gove, Phys. Rev. **87**, 262 (1952).
St 52a P. H. Stelson and W. M. Preston, Phys. Rev. **86**, 807 (1952).
St 53 P. Stähelin and P. Preiswerk, Nuovo cimento **10**, 1219 (1953).
St 53a P. Stähelin, Phys. Rev. **92**, 1076(L) (1953).
St 53b P. Stähelin, Helv. Phys. Acta **26**, 691 (1953).
St 54 P. H. Stelson and W. M. Preston, Phys. Rev. **95**, 974 (1954).
St 54a P. H. Stelson, Phys. Rev. **96**, 1584 (1954).
St 56 T. Stribel, Z. Naturforsch. **11a**, 166(L) (1956).
St 56a T. Stribel, Z. Naturforsch. **11a**, 254(L) (1956).
St 56b G. S. Stanford and G. F. Pieper, Phys. Rev. **103**, 637 (1956).
Su 53 Summers-Gill, Haslam, and Katz, Can. J. Phys. **31**, 70 (1953).
Su 55 A. D. Suttle Jr. and W. F. Libby, Anal. Chem. **27**, 921 (1955).
Sw 52 C. P. Swann and C. E. Mandeville, Phys. Rev. **87**, 215(A) (1952).
Sw 56 C. P. Swann and W. C. Porter, Bull. Am. Phys. Soc. Ser. II. **1**, 29 (1956).
Ta 46 R. Tangen, Det Kgl. Norske Videnskabers Selskabs Skrifter (1946).
Ta 53 Takemoto, Dazai, and Chiba, Phys. Rev. **91**, 1024(L) (1953).
Ta 54 G. E. Tauber and T.-Y. Wu, Phys. Rev. **93**, 295 (1954).
Ta 54a G. E. Tauber and T.-Y. Wu, Phys. Rev. **94**, 1307 (1954).
Ta 54b Taylor, Russell, and Cooper, Phys. Rev. **93**, 1056 (1954).
Ta 55 Taylor, Lönsjö, and Bonner, Phys. Rev. **100**, 174 (1955).
Ta 56 Taylor, Marshak, and Newson, Bull. Am. Phys. Soc. Ser. II, **1**, 62 (1956) plus verbal report to Nuclear Data Group.
Te 54 G. M. Temmer and N. P. Heydenburg, Phys. Rev. **96**, 426 (1954).
Te 54a Teener, Seagondollar, and Krone, Phys. Rev. **93**, 1035 (1954).
Te 56 G. M. Temmer and N. P. Heydenburg, Phys. Rev. **104**, 989 (1956).
Th 54 L. C. Thompson, Phys. Rev. **96**, 369 (1954).
Th 56 R. Thieberger and I. Talmi, Phys. Rev. **102**, 923(L) (1956).
Ti 51 H. K. Ticho, Phys. Rev. **84**, 847(L) (1951).
To 49 C. H. Townes and L. C. Aamodt, Phys. Rev. **76**, 691(L) (1949).
To 52 Toops, Sampson, and Steigert, Phys. Rev. **85**, 280 (1952).
To 53 B. J. Toppel and S. D. Bloom, Phys. Rev. **91**, 473(A) (1953).
To 55 Toller, Patterson, and Newson, Phys. Rev. **99**, 620(A) (1955).
To 55a Toller, Newson, and Merzbacher, Phys. Rev. **99**, 1625(A) (1955).
To 55b F. M. Tomnovec and C. S. Cook, Phys. Rev. **100**, 1254(A) (1955).
To 55c J. Tobailem, J. phys. radium **16**, 48 (1955).
To 57 Towle, Berenbaum, and Matthews, Proc. Phys. Soc. London **A70**, 84 (1957).
Tr 56 B. Trumpy and A. Graue, Physica **22**, 1155(A) (1956).
Tu 51 J. F. Turner and P. E. Cavanagh, Phil. Mag. **42**, 636 (1951).
Tu 53 O. H. Turner, Australian J. Phys. **6**, 380 (1953).
Tu 54 A. Turkevich and A. Samuels, Phys. Rev. **94**, 364 (1954).
Ty 54 H. Tyren and P.-A. Tove, Phys. Rev. **96**, 773 (1954).
Va 52 D. M. Van Patter and W. W. Buechner, Phys. Rev. **87**, 51 (1952).
Va 52a Van Patter, Sperduto, Endt, Buechner, and Enge, Phys. Rev. **85**, 142(L) (1952).
Va 52b Van Patter, Endt, Sperduto, and Buechner, Phys. Rev. **86**, 502 (1952).
Va 53 J. J. Van Loef (private communication).
Va 56 Van Patter, Swann, Porter, and Mandeville, Phys. Rev. **103**, 656 (1956).
Va 56a J. J. Van Loef and D. A. Lind, Phys. Rev. **101**, 103 (1956).
Va 56b J. Varma, Proc. Phys. Soc. (London) **A69**, 641(L) (1956).
Va 56c J. Varma and W. Jack, Physica **22**, 1139(A) (1956).
Va 56d Van der Leun, Endt, Kluyver, and Vrenken, Physica **22**, 1223 (1956).
Va 56e I. J. Van Heerden and D. J. Prowse, Phil. Mag. **1**, 967(L) (1956).
Va 56f C. van der Leun and P. M. Endt, Physica **22**, 1234(L) (1956).
Va 57 C. van der Leun and P. M. Endt (to be published.)
Va 57a Van Patter, Rothman, Porter, and Mandeville, Phys. Rev. **107**, 171 (1957).
Va 57b R. Van Lieshout and R. W. Hayward (private communication).

Va 57c Van Patter, Rothman, Porter, and Mandeville, Bull. Am. Phys. Soc. Ser. II, 2, 60 (1957).
Va 57d Van Patter, Porter, and Rothman, Phys. Rev. **106**, 1016 (1957).
Ve 56 S. H. Vegors and P. Axel, Phys. Rev. **101**, 1067 (1956).
Ve 56a S. H. Vegors and R. B. Duffield, Bull. Am. Phys. Soc. Ser. II, 1, 206 (1956).
Ve 57 Verbinski, Hurlimann, Stephens, and Winhold, Bull. Am. Phys. Soc. Ser. II, 2, 68 (1957).
Vo 56 P. Von Herrmann and G. F. Pieper, Bull. Am. Phys. Soc. Ser. II, 1, 281 (1956).
Vo 57 Vorona, Olness, Haeberli, and Lewis, Bull. Am. Phys. Soc. Ser. II, 2, 34 (1957).
Wa 37 H. Walke, Phys. Rev. **52**, 663(L) (1937).
Wa 39 Y. Watase and J. Itoh, Proc. Phys.-Math. Soc. Japan **21**, 626 (1939).
Wa 41 Y. Watase, Proc. Phys.-Math. Soc. Japan **23**, 618 (1941).
Wa 50 Warshaw, Chen, and Appleton, Phys. Rev. **80**, 288(L) (1950).
Wa 52 A. H. Wapstra, Phys. Rev. **86**, 561(L) (1952).
Wa 53 A. H. Wapstra and A. L. Veenendaal, Phys. Rev. **91**, 426(L) (1953).
Wa 53a A. H. Wapstra, Arkiv Fysik **6**, 263 (1953).
Wa 54 N. S. Wall, Phys. Rev. **96**, 664 (1954).
Wa 55 Way, King, McGinnis, and Van Lieshout, Nuclear Level Schemes ($A=40$–$A=92$), US-AEC Report TID-5300 (1955).
Wa 55a A. H. Wapstra, Physica **21**, 367 (1955).
Wa 55b A. H. Wapstra, Physica **21**, 385 (1955).
Wa 55c G. J. Wasserburg and R. J. Hayden, Phys. Rev. **97**, 86 (1955).
Wa 56 H. Wäffler and F. Heinrich, Physica **22**, 1146(A) (1956).
Wa 56a H. J. Watters, Phys. Rev. **103**, 1763 (1956).
We 51 Wentink, Koski, and Cohen, Phys. Rev. **81**, 948 (1951).
We 52 T. Westermark, Phys. Rev. **88**, 573 (1952).
We 54 P. Weinzierl, Z. Naturforsch. **9a**, 69 (1954).
We 56 Wetherill, Wasserburg, Aldrich, Tilton, and Hayden, Phys. Rev. **103**, 987 (1956).
We 56a J. B. Weddell, Phys. Rev. **104**, 1069 (1956).
Wh 39 White, Delsasso, Fox, and Creutz, Phys. Rev. **56**, 512 (1939).
Wh 41 White, Creutz, Delsasso, and Wilson, Phys. Rev. **59**, 63 (1941).
Wh 53 R. L. White and C. H. Townes, Phys. Rev. **92**, 1256 (1953).
Wi 41 C. M. Witcher, Phys. Rev. **60**, 32 (1941).
Wi 51 D. H. Wilkinson and J. H. Carver, Phys. Rev. **83**, 466(L) (1951).
Wi 52 Willard, Kington, and Bair, Phys. Rev. **86**, 259(L) (1952).
Wi 54 Williams, McCall, and Gutowsky, Phys. Rev. **93**, 1428 (1954).
Wi 55 R. M. Williamson and S. A. Cox, Phys. Rev. **99**, 644(A) (1955).
Wi 55a A. Winther, J. phys. radium **16**, 562 (1955).
Wi 56 Willard, Bair, Cohn, and Kington, Bull. Am. Phys. Soc. Ser. II, 1, 264 (1956).
Wi 56a D. H. Wilkinson, Phil. Mag. **1**, 1031 (1956).
Wo 50 D. C. Worth, Phys. Rev. **78**, 378 (1950).
Wo 54 C. Wong, Phys. Rev. **95**, 761 (1954).
Wo 54a M. Wolfsberg, Phys. Rev. **96**, 1712(L) (1954).
Wo 56 E. A. Wolf, Phil. Mag. **1**, 102(L) (1956).
Wo 56a J. L. Wolfson, Can. J. Phys. **34**, 256 (1956).
Wr 53 B. T. Wright, Phys. Rev. **90**, 159(L) (1953).
Wu 49 Wu, Townes, and Feldman, Phys. Rev. **76**, 692(L) (1949).
Wu 50 C. S. Wu, Revs. Modern Phys. **22**, 386 (1950).
Ye 56 P. F. Yergin, Phys. Rev. **104**, 1340 (1956).
Ze 52 Zeldes, Ketelle, Brosi, Fultz, and Hibbs, Phys. Rev. **86**, 811(L) (1952).
Zu 50 A. Zucker and W. W. Watson, Phys. Rev. **80**, 966 (1950).
Zw 54 P. F. Zweifel, Phys. Rev. **96**, 1572 (1954) plus Erratum in Phys. Rev. **98**, 1870 (1955).

Nuclear Disintegration Energies. II.*†

D. M. VAN PATTER, *Bartol Research Foundation, Swarthmore, Pennsylvania*

AND

WARD WHALING, *California Institute of Technology, Pasadena, California*

I. INTRODUCTION

THIS supplement is intended to augment the compilation, "Nuclear Disintegration Energies,"[1] which included measurements of reaction energies received up to May, 1954. The data contained in the original compilation, together with other data pertaining to mass differences[2] have been of use in the preparation of tables of masses [3–5] and mass links,[6] and for the comparison of masses or mass differences obtained from nuclear reaction energies and mass-spectroscopic measurements.[3,7]

Since publication of the original compilation, there has been a considerable number of new measurements of nuclear reaction energies, e.g., measurements of some eighty previously undetermined reaction energies have been reported for $A<50$. The systematic program of precise Q-value determinations by the group at Massachusetts Institute of Technology headed by W. W. Buechner has been extended up to the mass region $A=32-60$. Together with data concerning beta-decay energies, these measurements have made possible the calculation of masses in the region from S^{32} to Ti^{45} from nuclear reaction energies only.[8]

In his systematic examination of masses and mass differences for $A<202$, Wapstra[3] pointed out several instances of Q values listed in our original compilation[1] which should be assigned to transitions to excited states, rather than ground-state transitions. He also discovered discrepancies between nuclear reaction and mass-spectroscopic data, particularly in the region of the nickel isotopes. On the basis of his analysis, he also reassigned some of the gamma rays observed from neutron capture by Kinsey and Bartholomew to other isotopes. A similar detailed analysis of (n,γ) and (γ,n) transitions has been made by Way *et al.*[6,9] in order to determine neutron binding energies in the region of $A=40-92$. More recently, Quinsenberry, Scolman, and Nier,[10] on the basis of their new mass measurements in the region Fe to Zn, have reassigned several neutron capture gamma rays to new isotopes, particularly for the nickel and zinc isotopes. It can be seen from the above brief summary that a considerable alteration of the nuclear reaction energies contained in the original compilation is now necessary.

This supplement contains new measurements of nuclear reaction energies available to the authors up to February 1, 1957. In addition, any measured values listed in the original compilation which have been revised or reassigned since its publication have been included.

II. ARRANGEMENT OF TABLE I

The arrangement of Table I of this supplement is the same as in the original compilation.[1] In Columns 1 and 2, the reactions and energy determinations are listed. In the case of a reaction shown as (n,γ,γ), the ground-state transition was not observed, and the reaction energy shown was determined from the summing of two gamma-ray energies. Some of the experimental values, designated by the superscript "a," do not appear explicitly in the reference cited, and have been calculated by the present authors from the pertinent experimental data. If it was necessary to use masses in such a calculation, those given by Wapstra[3] have been used. When there is doubt that a reported Q value represents the ground-state transition, it has been enclosed in parentheses.

In Columns 3 and 4, the method and energy standards used are listed, if stated by the authors. In Column 5, the most recent reference for each measured value is

* The contribution of D. M. Van Patter was supported in part by the U. S. Air Force, through the Office of Scientific Research of the Air Research and Development Command.

† This paper, and the three which follow, by H. E. Duckworth, L. J. Lidofsky, and by F. Asaro and I. Perlman, were prepared at the suggestion of the Subcommittee on Nuclear Constants of the Committee on Nuclear Science of the National Research Council. This subcommittee consists of Ward Whaling, *chairman*, and F. A. Selove, G. A. Bartholomew, H. E. Duckworth, I. Perlman, W. H. Sullivan, and D. M. Van Patter, together with L. Lidofsky as consultant.

Reprints of these four papers as a group may be obtained from the Publications Office, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

[1] D. M. Van Patter and W. Whaling, Revs. Modern Phys. **26**, 402 (1954).

[2] R. W. King, Revs. Modern Phys. **26**, 327 (1954); Geschwind, Gunther-Mohr, and Townes, Revs. Modern Phys. **26**, 444 (1954); F. Asaro and I. Perlman, Revs. Modern Phys. **26**, 456 (1954); Duckworth, Hogg, and Pennington, Revs. Modern Phys. **26**, 463 (1954).

[3] A. H. Wapstra, Physica **21**, 367 (1955); *ibid.* **21**, 385 (1955).

[4] J. R. Huizenga, Physica **21**, 410 (1955).

[5] Mattauch, Waldmann, Bieri, and Everling, Z. Naturforsch. **11A**, 525 (1956).

[6] Way, King, McGinnis, and van Lieshout, "Nuclear Level Schemes," TID-5300 (USAEC September 1955).

[7] M. E. Kettner, Phys. Rev. **102**, 1065 (1956); Scolman, Quisenberry, and Nier, Phys. Rev. **102**, 1076 (1956).

[8] Endt, Buechner, Braams, Paris, and Sperduto, Phys. Rev. **105**, 1002 (1957).

[9] The authors wish to acknowledge the valuable contributions of the Nuclear Data Group under the leadership of K. Way. Their compilation "Nuclear Level Schemes," for $40\leq A\leq 92$, and their tables of ground-state Q values in *Nuclear Science Abstracts* have aided considerably the preparation of this present supplement.

[10] Quisenberry, Scolman, and Nier, Phys. Rev. **104**, 461 (1956).

TABLE I. Nuclear reaction energies.

Reaction	Measured Q value (Mev)	Method	Calibration Energy standard	Reference	Average Q value (Mev)
$D^2(d,p)T^3$	4.044 ± 0.005	el spec	$Li^7(p,n)Be^7$	Do 56a	4.038 ± 0.005
$D^2(d,n)He^3$	3.276 ± 0.024	ph pl	Ro 51c	Su 55	3.267 ± 0.007
	3.271 ± 0.011	el spec	$Li^7(p,n)Be^7$	Do 56a	
$He^4(d,p)He^5$	-3.3 ± 0.1	ph pl		Fr 54a	-3.3 ± 0.1
$Li^6(n,\gamma)Li^7$	7.26 ± 0.05	pr spec		Ba 56	
$Li^6(p,d)Li^5$	-3.0 ± 0.15	scint spec		Li 55	preliminary
$Li^6(p,\gamma)Be^7$	5.66 ± 0.03	scint spec	$F^{19}(p,\alpha\gamma)O^{16}$	Wa 56	
$Li^6(d,He^3)He^5$	0.91 ± 0.09	mag spec	Po α	Le 55	
$Li^6(t,\alpha)He^5$	15.15 ± 0.04	pulse ht	Th C' α	Cr 56a	
$Li^6(t,d)Li^7$	0.986 ± 0.007	mag spec	$Li^6(p,\alpha)He^3$	Pe 52, Al 54*	$0.986\pm0.007^+$
$Li^6(t,p)Li^8$	0.790 ± 0.011	mag spec	$Li^6(p,\alpha)$, $Li^6(t,d)Li^{7*}$	Pe 52, Al 54*	
$Li^6(He^3,n)B^8$	-1.976 ± 0.006	threshold		Du 56	
$Li^7(p,n)Be^7$	$-1.645_2\pm0.001$	threshold	$Mg^{24}(p,p')$, $Na^{24}\gamma$, $Au^{198}\gamma$	Jo 54	$-1.644_9\pm0.000_4$
	$-1.643_7\pm0.001$	threshold	$Mg^{24}(p,p')$, $Na^{24}\gamma$, $Au^{198}\gamma$	Jo 54	
$Li^7(d,\alpha)He^5$	13.719	mag spec	$Li^6(d,p)Li^7$	Kh 55	
	14.26 ± 0.09	mag spec	Po α	Le 55	
$Li^7(d,p)Li^8$	-0.183	mag spec	$Li^6(d,p)Li^7$	Kh 55	-0.192 ± 0.001
$Li^7(\alpha,n)B^{10}$	-2.82 ± 0.10	pulse ht threshold	$D^2(d,n)$, $B^{10}(\alpha,p)$	Ro 56	
$Li^7(Li^7,p)B^{13}$	5.97 ± 0.05	range	$B^{10}(d,p)B^{11*}$	No 57	
$Be^8\rightarrow 2He^4$	0.090 ± 0.005	ang corr		Tr 55	$0.094_1\pm0.000_7$
	$0.093_9\pm0.000_8$	mag spec		Fo 56	
$Be^9(\gamma,n)Be^8$	-1.664 ± 0.004	threshold		Co 56	$-1.665\pm0.001_4$
$Be^9(n,\gamma)Be^{10}$	6.80	compt spec		Gr 55b	6.816 ± 0.006
$Be^9(p,pn)Be^8$	-1.664 ± 0.005	mag spec	Po α	Bo 56	
$Be^9(p,n)B^9$	-1.853 ± 0.003	threshold	$Li^7(p,n)Be^7$	Ma 55b	-1.852 ± 0.002
$Be^9(d,p)Be^{10}$	4.586 ± 0.009	mag spec	Po α	Ju 54	4.587 ± 0.005
$Be^9(d,n)B^{10}$	4.43 ± 0.08	ph pl	La 47	Pr 52, Pr 53*	4.35 ± 0.02
	4.28 ± 0.10	ph pl		Re 54	
	4.54 ± 0.06[c]	ph pl		Gr 55	
$Be^9(Li^7,p)C^{15}$	9.05 ± 0.05	range	$Li^7(Li^7,p)B^{12}$	No 57	
$B^{10}(n,t)Be^8$	(0.35 ± 0.20)	pulse ht	$D^2(d,n)$, $B^{10}(n,\alpha)Li^{7*}$	Ja 55	
$B^{10}(n,\gamma)B^{11}$	11.43 ± 0.04	pr spec	absolute	Ba 56	
$B^{10}(p,n)C^{10}$	-4.37 ± 0.05	threshold		Co 55	-4.37 ± 0.05
$B^{10}(p,\gamma)C^{11}$	8.81 ± 0.10	scint spec	$F^{19}(p,\alpha\gamma)$, $C^{13}(p,\gamma)$	Ch 56	8.81 ± 0.10
$B^{10}(d,\alpha)Be^8$	17.829 ± 0.010	mag spec	absolute	El 54	17.829 ± 0.010
$B^{10}(d,t)B^9$	-2.187 ± 0.010	mag spec	Po α	Bo 56	
$B^{10}(d,p)B^{11}$	9.227 ± 0.006	mag spec	absolute	El 54	9.229 ± 0.005
$B^{10}(He^2,n)N^{12}$	1.46 ± 0.06	ph pl		Aj 57	preliminary
$B^{10}(\alpha,d)C^{12}$	1.36 ± 0.09	scint spec	$Al^{27}(\alpha,p)Si^{30}$	Pi 56	1.341 ± 0.002
	1.341 ± 0.002	el spec	$Li^7(p,n)Be^7$	Do 56a	
$B^{10}(\alpha,p)C^{13}$	4.064 ± 0.012	mag spec	Po α	Fa 55	4.064 ± 0.012
	4.08 ± 0.03	scint spec	$Al^{27}(\alpha,p)Si^{30}$	Pi 56	
	4.10 ± 0.03	ph pl	Ro 51c	Pa 56	
$B^{11}(p,n)C^{11}$	$-2.83^{+0.08}_{-0.05}$	ph pl	Ro 51c	Aj 56	-2.762 ± 0.003
$B^{11}(d,\alpha)Be^9$	8.029 ± 0.005	mag spec	absolute	El 54	8.024 ± 0.004
	8.015 ± 0.010	mag spec	Po α	Bo 56	
$B^{11}(d,p)B^{12}$	1.110	spec		Kh 54	
$B^{11}(d,n)C^{12}$	13.81	ph pl		Ih 55	13.8 ± 0.1
$B^{11}(\alpha,p)C^{14}$	0.788 ± 0.017	mag spec	Po α	Fa 55	0.788 ± 0.017
$B^{11}(\alpha,n)N^{14}$	0.0 ± 0.3	pulse ht	$D^2(d,n)He^3$, $B^{10}(\alpha,p)C^{13}$	Qu 56	$0.27\pm0.06^+$
$C^{12}(p,\alpha)B^9$	-7.58 ± 0.10	range	Be 50d, Bi 54a	Re 55	
$C^{12}(d,\alpha)B^{10}$	-1.39 ± 0.02	range	El 51c	El 57	
$C^{12}(d,p)C^{13}$	2.717 ± 0.010	mag spec	Po α	Sp 54	$2.721\pm0.002^+$
	2.720 ± 0.003	mag spec	absolute	El 54	

TABLE I.—*Continued.*

Reaction	Measured Q value (Mev)	Method	Calibration energy standard	Reference	Average Q value (Mev)
$C^{12}(He^3,n)O^{14}$	$-1.147\pm0.002_4$[a]	threshold		Bu 56a	-1.152 ± 0.005
	$-1.158_5\pm0.003$	threshold	$Li^7(p,n)Be^7$, $Li^7(\alpha,\gamma)B^{11}$	Br 56a	
$C^{13}(p,d)C^{12}$	-2.718	mag spec	Po α	Sp 56a	$-2.721\pm0.002^+$
$C^{13}(d,p)C^{14}$	5.942 ± 0.011	mag spec	Po α	Sp 54	5.943 ± 0.003
	5.953 ± 0.010	mag spec	ThC α	Ah 54d	
$C^{13}(d,n)N^{14}$	5.40 ± 0.10	ph pl	Ri 51	Be 52	5.35 ± 0.03
	5.41 ± 0.06	ph pl		Gr 55	
	5.325 ± 0.04	ph pl		Bi 55	
$C^{14}(p,n)N^{14}$	$-0.626_4\pm0.000_5$	threshold	$Li^7(p,n)Be^7$	Sa 56	$-0.626_4\pm0.000_5^+$
$C^{14}(d,\alpha)B^{12}$	$0.362\pm0.001_5$	el spec	$Li^7(p,n)Be^7$	Do 56a	
$C^{14}(d,p)C^{15}$	-1.007 ± 0.001	el spec	$Li^7(p,n)Be^7$	Do 56a	
$C^{14}(\alpha,n)O^{17}$	-1.820 ± 0.002	threshold	$Li^7(p,n)Be^7$	Sa 56	
$N^{14}(n,p)C^{14}$	0.609 ± 0.005[c]	pulse ht		Is 50	$0.626_4\pm0.000_5^+$
$N^{14}(d,n)O^{15}$	5.21 ± 0.07	ph pl		No 56	5.14 ± 0.03
$N^{14}(\alpha,p)O^{17}$	-1.16	ph pl		Hj 53, Hj 53a	-1.16 ± 0.04
$N^{14}(\alpha,n)F^{17}$	-4.76 ± 0.07	threshold, pulse ht	Be 49	Do 56	-4.76 ± 0.07
$N^{15}(p,n)O^{15}$	-3.539 ± 0.008	threshold	(p,n) threshs, (p,γ) resonances	Ki 55	-3.5432 ± 0.0015
	-3.5432 ± 0.0015	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Li 57	preliminary
$N^{15}(d,p)N^{16}$	0.286	mag spec		Wh 55	
$O^{16}(d,p)O^{17}$	1.915 ± 0.010	mag spec	Po α	Sp 54	1.918 ± 0.004
$O^{16}(d,n)F^{17}$	-1.622 ± 0.004	threshold	$Li^7(p,n)Be^7$	Bo 51, Ma 55a*	-1.623 ± 0.003
	-1.626 ± 0.004	threshold	(p,n) thresholds	Ma 55a	
$O^{17}(d,\alpha)N^{15}$	9.807 ± 0.012	mag spec	$F^{19}(d,\alpha)O^{17}$, $N^{14}(d,\alpha)C^{12}$	Pa 54a	
$O^{17}(d,p)O^{18}$	5.821 ± 0.010	mag spec	ThC α	Ah 54c, Ah 54d	
$O^{18}(p,\alpha)N^{15}$	3.967 ± 0.009	mag spec	$F^{19}(p,\alpha\gamma)O^{16}$, $N^{15}(p,\alpha)C^{12}$	Mi 54, Ah 54d*	3.967 ± 0.009
$O^{18}(p,n)F^{18}$	-2.447 ± 0.010	threshold	$Li^7(p,n)Be^7$	Ma 56a	-2.452 ± 0.004
$O^{18}(d,\alpha)N^{16}$	4.237 ± 0.009	mag spec	$C^{12}(d,p)C^{13}$, $N^{14}(d,p)N^{15}$*	Pa 55a	
$O^{18}(d,p)O^{19}$	1.730 ± 0.008	mag spec	$B^{10}(d,p)B^{11}$*, $F^{19}(p,\alpha\gamma)O^{16}$	Ah 54b, Mi 54b	1.732 ± 0.005
	1.732 ± 0.008	mag spec		Th 54	
	1.735 ± 0.008	mag spec		Ho 55	
$F^{19}(n,\gamma)F^{20}$	6.599 ± 0.011	pr spec	absolute	Ca 56	6.599 ± 0.011
$F^{19}(p,\alpha)O^{16}$	8.110 ± 0.010	mag spec	Po α	Sq 56	8.114 ± 0.007
$F^{19}(p,n)Ne^{19}$	-4.029 ± 0.008	threshold	$F^{19}(p,\alpha\gamma)O^{16}$	Wi 52, Ki 55*	-4.025 ± 0.004
	-4.027 ± 0.008	threshold	(p,n)threshs, (p,γ)resonances	Ki 55	
	-4.022 ± 0.005	threshold	$Li^7(p,n)Be^7$	Ma 55b	
$F^{19}(d,p)F^{20}$	4.38 ± 0.03	range	El 51c	El 56	4.373 ± 0.007
$F^{19}(d,t)F^{18}$	-4.17 ± 0.02	range	El 51c	El 57	
$F^{19}(t,p)F^{21}$	6.200 ± 0.025	mag spec	$O^{16}(t,\alpha)N^{15}$	Ja 56	6.200 ± 0.025
	6.03 ± 0.1[a]	ph pl		Bi 55a	preliminary
$F^{19}(\alpha,p)Ne^{22}$	1.673 ± 0.011	mag spec	Po α	Fa 55	1.673 ± 0.011
$F^{19}(\alpha,n)Na^{22}$	-2.0 ± 0.2	pulse ht	$D^2(d,n)He^3$, $B^{10}(\alpha,p)C^{13}$	Qu 56	-2.0 ± 0.2
$Ne^{20}(n,\alpha)O^{17}$	-0.70 ± 0.02	pulse ht		Fl 53	-0.71 ± 0.02
$Ne^{20}(d,\alpha)F^{18}$	2.810 ± 0.009	mag spec	$F^{19}(p,\alpha\gamma)O^{16}$, $O^{16}(d,p)O^{17}$	Mi 54a, Mi 56*	2.810 ± 0.009
$Ne^{22}(d,p)Ne^{23}$	2.968 ± 0.008	mag spec	$C^{12}(d,p)C^{13}$	Ah 54a	2.966 ± 0.005
$Na^{23}(n,\gamma,\gamma)Na^{24}$	6.96 ± 0.03	compt spec		Gr 55b	
$Na^{23}(p,\alpha)Ne^{20}$	2.370 ± 0.008	mag spec	Po α	Bu 56c	2.377 ± 0.003
$Na^{23}(p,n)Mg^{23}$	-4.849 ± 0.010	threshold	$F^{19}(p,\alpha\gamma)O^{16}$	Wi 52a, Ki 55*	-4.845 ± 0.007
	-4.841 ± 0.010	threshold	(p,n)threshs, (p,γ)resonances	Ki 55	
$Na^{23}(\alpha,n)Al^{26}$	-2.9 ± 0.2	pulse ht	Be 49	Do 56	
	$\geq -2.970\pm0.004$	threshold	$Li^7(p,n)Be^7$	Bu 56b	preliminary
$Mg^{24}(p,\gamma)Al^{25}$	2.26 ± 0.03[a]	scint spec	$F^{19}(p,\alpha\gamma)O^{16}$, (p,γ)reactions	Gr 55a	$2.28\pm0.01_5$
	2.29 ± 0.02[a]	scint spec	$Pr^{144}\gamma$, $Cs^{137}\gamma$, $Na^{22}\gamma$	Cr 56	
	2.29 ± 0.02[a]	scint spec	$Zn^{65}\gamma$, $Co^{60}\gamma$, $ThC''\gamma$	Ag 56	
$Mg^{24}(d,\alpha)Na^{22}$	1.953 ± 0.012	mag spec	Po α	Br 55, Br 55a*	preliminary
$Mg^{24}(d,p)Mg^{25}$	5.02 ± 0.02[c]	spec		Kh 53	5.097 ± 0.007

TABLE I.—*Continued.*

Reaction	Measured Q value (Mev)	Method	Calibration energy standard	Reference	Average Q value (Mev)
$Mg^{25}(\gamma,p)Na^{24}$	−12.1	threshold		Ka 54	
$Mg^{25}(\gamma,n)Mg^{24}$	−7.2	threshold		Ka 54	
$Mg^{25}(n,\gamma)Mg^{26}$	11.086±0.025	pr spec	absolute	Ca 56	
$Mg^{25}(p,\alpha)Na^{22}$	−3.15	mag spec	Po α	Br 55a	preliminary
$Mg^{25}(p,n)Al^{25}$	−5.084±0.024	threshold	(p,n)threshs, (p,γ)resonances	Ki 55	
$Mg^{25}(p,\gamma)Al^{26}$	6.35±0.08	scint spec	$F^{19}(p,\alpha\gamma)$, $Be^{9}(\alpha,n\gamma)$, $C^{13}(p,\gamma)$	Kl 54	
$Mg^{25}(\alpha,p)Al^{28}$	−1.29±0.04	pulse ht	$Al^{27}(p,\alpha)Mg^{24}$	Gr 57	
$Mg^{26}(\gamma,p)Na^{25}$	−14.3	threshold		Ka 54	
$Mg^{26}(\gamma,n)Mg^{24}$	−11.1	threshold		Ka 54	
$Mg^{26}(p,n)Al^{26}$	−4.83±0.1[a]	threshold		Sc 54 } En 54, Ka 55	−4.778±0.015
	−4.778±0.015[a]	threshold	(p,n)threshs, (p,γ)resonances	Ki 55 }	
$Mg^{26}(\alpha,p)Al^{29}$	−2.90±0.04	pulse ht	$Al^{27}(p,\alpha)Mg^{24}$	Gr 57	
$Al^{27}(\gamma,n)Al^{26}$	−13.4±0.2	threshold	$Cu^{63}(\gamma,n)Cu^{62}$	Ha 54	−13.1+0.3
$Al^{27}(p,\alpha)Mg^{24}$	1.61±0.04	ph pl		Gr 54	1.595±0.002[+]
	1.596±0.006	mag spec	$Li^{7}(p,n)Be^{7}$, Po α	Va 57	
$Al^{27}(p,n)Si^{27}$	−5.581±0.010	threshold	$F^{19}(p,\alpha\gamma)O^{16}$	Ki 53d, Ki 55*	−5.593±0.009
	−5.584±0.010	threshold	(p,n)threshs, (p,γ)resonances	Ki 55	
	−5.607±0.008	threshold	$Li^{7}(p,n)Be^{7}$, $E_T=1.8811$	Ma 55b	
$Al^{27}(d,p)Al^{28}$	5.475	spec		Kh 54	5.498±0.007
	5.502±0.010	mag spec	Po α	Bu 56	
$Al^{27}(\alpha,p)Si^{30}$	2.38±0.03	scint spec		Ha 56	2.38±0.03
		pulse ht			
		ph pl			
$Al^{27}(\alpha,n)P^{30}$	$\geq$ −2.662±0.004	threshold	$Li^{7}(p,n)Be^{7}$	Bu 56b	preliminary
$Si^{28}(n,\gamma)Si^{29}$	8.482±0.015	compt spec		Ad 56a	8.471±0.007
$Si^{28}(d,\alpha)Al^{26}$	1.416±0.008	mag spec	Po α	Br 54c	
$Si^{29}(\gamma,p)Al^{28}$	−12.3	threshold		Ka 54	
$Si^{29}(\gamma,n)Si^{28}$	−8.5	threshold		Ka 54	
$Si^{29}(n,\gamma)Si^{30}$	10.59±0.03	compt spec		Ad 56a	10.600±0.010
$Si^{29}(p,\gamma)P^{30}$	5.55±0.06	scint spec	$Be^{9}(\alpha,n\gamma)C^{12}$, $F^{19}(p,\alpha\gamma)O^{16}$	En 54a	
$Si^{30}(\gamma,p)Al^{29}$	−12.9	threshold		Ka 54	
$Si^{30}(\gamma,n)Si^{29}$	−10.6	threshold		Ka 54	
$P^{31}(\gamma,n)P^{30}$	−12.33±0.05	threshold	(γ,n)thresholds	Ba 55a	−12.32±0.05
$P^{31}(p,\alpha)Si^{28}$	1.911±0.005	mag spec	$Li^{7}(p,n)$, $Al^{27}(p,\alpha)$, Po α	Va 56	1.910±0.004
	1.909±0.010	mag spec	Po α	En 57	
$P^{31}(p,n)S^{31}$	−6.06±0.2	ph pl	Gi 54	Ru 56	
$P^{31}(d,n)S^{32}$	6.63±0.08	ph pl	El 51d	El 52, El 55*	6.63±0.08
$P^{31}(\alpha,p)S^{34}$	0.7±0.1	scint spec	$Al^{27}(\alpha,p)Si^{30}$	St 56	0.7±0.1
$P^{31}(\alpha,n)Cl^{34}$	−5.7±0.2	threshold	$D^{2}(d,n)He^{3}$, $B^{10}(\alpha,p)C^{13}$	Qu 56	−5.7±0.2
$S^{32}(n,\gamma)S^{33}$	8.63±0.04	compt spec		Gr 55b	8.64±0.02
$S^{32}(p,\gamma)Cl^{33}$	2.285±0.012	scint spec	$Na^{24}\gamma$	Va 56a	
$S^{32}(d,\alpha)P^{30}$	4.831±0.013	mag spec		Le 56	
$S^{32}(d,p)S^{33}$	6.408±0.020	mag spec		Le 56	6.419±0.010
$S^{32}(\alpha,p)Cl^{35}$	−2.3	scint spec	$Al^{27}(\alpha,p)Si^{30}$	Pi 55	−1.861±0.004[+]
$S^{34}(p,n)Cl^{34}$	(−6.1)	ph pl		Aj 55	preliminary
$S^{34}(d,\alpha)P^{32}$	5.04±0.02	mag spec		Le 56	
$Cl^{35}(\gamma,n)Cl^{34}$	−12.35±0.035	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$threshs	De 55	
$Cl^{35}(n,\gamma)Cl^{36}$	8.55±0.04	compt spec		Gr 55b	8.56±0.02
$Cl^{35}(p,\alpha)S^{32}$	1.865±0.015	mag spec		Al 55	1.861±0.004[+]
	1.863±0.008	mag spec	Po α	En 56	
	1.860±0.005	mag spec	$Li^{7}(p,n)Be^{7}$, Po α	Va 56, Va 57*	
$Cl^{35}(d,\alpha)S^{33}$	8.277±0.010	mag spec	Po α	Pa 55	8.277±0.010
$Cl^{35}(d,p)Cl^{36}$	6.354±0.008	mag spec	Po α	Pa 55	6.354±0.008
$Cl^{37}(p,\alpha)S^{34}$	3.015±0.015	mag spec		Al 55	3.026±0.005
	3.026±0.008	mag spec	Po α	En 56	
	3.028±0.006	mag spec	$Li^{7}(p,n)Be^{7}$, Po α	Va 56, Va 57*	

TABLE I.—*Continued.*

Reaction	Measured Q value (Mev)	Method	Calibration energy standard	Reference	Average Q value (Mev)
$Cl^{37}(d,\alpha)S^{35}$	7.783±0.012	mag spec	Po α	Pa 55	
$Cl^{37}(d,p)Cl^{38}$	3.877±0.008	mag spec	Po α	Pa 55	3.877±0.008
$A^{36}(\alpha,p)K^{39}$	−1.28±0.03	ph pl	$N^{14}(\alpha,p)O^{17}$	Sc 56a	−1.283±0.008⁺
$A^{40}(\gamma,n)A^{39}$	−9.85±0.15	threshold	C^{12}, Mn^{55}, $Bi^{209}(\gamma,n)$ threshs	Ha 54a	
$A^{40}(n,\alpha)S^{37}$	−2.5±0.1	pulse ht	Po α	Be 55	
$A^{40}(\alpha,p)K^{43}$	−3.36±0.03	ph pl	$N^{13}(\alpha,p)O^{17}$	Sc 56a	preliminary
$K^{39}(\gamma,n)K^{38}$	−13.00	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$ threshs	De 55	
$K^{39}(n,\gamma,\gamma)K^{40}$	7.795±0.010	compt spec		Ad 56, Bu 53	7.791±0.006
$K^{39}(n,\alpha)Cl^{36}$	1.25±0.20	scint spec	Po α, $K^{39}(n,p)A^{39}$	Sc 56	
$K^{39}(p,\alpha)A^{36}$	1.267±0.020 1.286±0.008	mag spec mag spec	 Po α	Al 55 Sp 55, Sp 56a*	1.283±0.008⁺
$K^{39}(\alpha,p)Ca^{42}$	−0.19±0.07	range		Sc 55	−0.118±0.007⁺
$K^{41}(p,\alpha)A^{38}$	4.002±0.015	mag spec	mag spec	Al 55	
$K^{41}(\alpha,p)Ca^{44}$	0.98±0.10	range		Sc 55	1.057±0.010⁺
$Ca^{40}(n,\gamma,\gamma)Ca^{41}$	8.37±0.03 8.350±0.017	pr spec compt spec	absolute	Ki 52, Br 56b Ad 56	8.355±0.014
$Ca^{40}(d,\alpha)K^{38}$	4.650±0.010	mag spec	Po α	Br 56	
$Ca^{40}(d,p)Ca^{41}$	6.140±0.009	mag spec	Po α	Br 54, Br 56b*	6.140±0.009
$Ca^{40}(d,n)Sc^{14}$	−0.60±0.05	range		Pl 55	preliminary
$Ca^{42}(p,\alpha)K^{39}$	0.118±0.007	mag spec	Po α	Br 56	0.118±0.007⁺
$Ca^{42}(d,p)Ca^{43}$	5.711±0.010	mag spec	Po α	Br 54b, Br 56*	
$Ca^{43}(p,\alpha)K^{40}$	−0.014±0.008	mag spec	Po α	Br 56	
$Ca^{43}(d,p)Ca^{44}$	9.07±0.07 8.913±0.014	range mag spec	 Po α	Sc 55 Br 56	8.913±0.014
$Ca^{44}(p,\alpha)K^{41}$	−1.057±0.010	mag spec	Po α	Br 56	−1.057±0.010⁺
$Ca^{44}(d,p)Ca^{45}$	5.188±0.010	mag spec	Po α	Br 54b, Br 56*	
$Ca^{48}(d,p)Ca^{49}$	2.916±0.006	mag spec	Po α	Br 56	2.916±0.006
$Sc^{45}(p,n)Ti^{45}$	−2.844±0.004	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Br 55b	
$Ti^{47}(n,\gamma,\gamma)Ti^{48}$	(11.609±0.02) [10.619 γ transition (Ba 56a) to 0.990±0.015 state (He 55)] (11.51±0.05)			Ba 56a Ki 53, Wa 55	
$Ti^{47}(d,p,\gamma)Ti^{48}$	9.13±0.05			Pi 52a, Wa 55	9.13±0.05
$Ti^{48}(n,\gamma,\gamma)Ti^{49}$	8.141±0.008 [6.756±0.006 γ transition (Ki 53) to 1.385±0.005 state (Mo 54)] 8.14±0.02 8.153±0.010			Ba 56a Ki 53, Wa 55 Ad 56a	8.145±0.006
$Ti^{49}(n,\gamma,\gamma)Ti^{50}$	10.97±0.07			Ki 53, Wa 55	
$V^{51}(p,\alpha)Ti^{48}$	1.161±0.010	mag spec	Po α	Bu 55	
$V^{51}(p,n)Cr^{51}$	−1.535±0.001 −1.536	threshold threshold	$Li^7(p,n)Be^7$ $Li^7(p,n)Be^7$, $E_T=1.8811$	Gi 55 Ma 56	−1.535±0.001
$Cr^{50}(n,\gamma,\gamma)Cr^{51}$	9.25±0.01 [8.499±0.007 γ transition (Ki 53) to 0.750±0.011 state (He 55)]			Ki 53, Wa 55	
$Cr^{52}(d,p)Cr^{53}$	5.74	range Al		El 56a	preliminary
$Cr^{53}(d,p)Cr^{54}$	7.55	range Al		El 56a	preliminary
$Cr^{54}(p,n)Mn^{54}$	−2.162±0.005	threshold	$Li^7(p,n)Be^7$	Lo 52	
$Mn^{55}(p,n)Fe^{55}$	−1.015±0.003	threshold		Jo 56	−1.014±0.003
$Fe^{54}(\gamma,n)Fe^{53}$	(−13.7±0.2) (−13.65±0.05) −11.90±0.07	threshold threshold threshold	 C^{12}, O^{16}, Cu^{63}, $Ag^{109}(\gamma,n)$ thresholds C^{12}, N^{14}, $O^{16}(\gamma,n)$ thresholds	Ka 51 Ba 55 De 55	
$Fe^{54}(n,\gamma)Fe^{55}$	9.295±0.015	compt spec		Ad 56a	9.297±0.006
$Fe^{54}(d,p)Fe^{55}$	7.073	mag spec	Po α	Sp 56	
$Fe^{56}(n,\gamma)Fe^{57}$	7.636±0.010	compt spec		Ad 56a	7.638±0.004
$Fe^{56}(d,p)Fe^{57}$	5.418 5.53	mag spec range	Po α	Sp 56 Mc 55	 preliminary
$Fe^{57}(d,p)Fe^{58}$	7.808 7.89	mag spec range	Po α	Sp 56 Mc 55	 preliminary

TABLE I.—*Continued.*

Reaction	Measured Q value (Mev)	Method	Calibration energy standard	Reference	Average Q value (Mev)
$Fe^{58}(d,p)Fe^{59}$	4.350	mag spec	Po α	Sp 56	
$Co^{59}(p,n)Ni^{59}$	-1.84	range ph pl		St 52	-1.858 ± 0.002
	-1.858 ± 0.004	threshold		Bu 56d	preliminary
	-1.862 ± 0.005^{b}	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Ch 57	
$Co^{59}(d,p)Co^{60}$	5.283 ± 0.008	mag spec	Po α	Fo 54	5.283 ± 0.008
$Ni^{58}(\gamma,n)Ni^{57}$	-12.0	threshold		Ka 51	
$Ni^{58}(n,\gamma)Ni^{59}$	8.996 ± 0.010	compt spec		Ad 56	8.997 ± 0.004
$Ni^{58}(p,\gamma)Cu^{59}$	3.42 ± 0.02	scint spec		Go 57a	
$Ni^{58}(d,p)Ni^{59}$	6.70 ± 0.1	range ph pl		Pr 54	6.74 ± 0.07
$Ni^{60}(n,\gamma)Ni^{61}$	(7.817 ± 0.008)	pr spec	absolute	Ki 53, Qu 56a	tgt isotope uncertain
	(7.825 ± 0.020)	compt spec		Ad 56, Qu 56a	tgt isotope uncertain
$Ni^{60}(p,\gamma)Cu^{61}$	4.81 ± 0.03	scint spec		Go 57a	preliminary
$Ni^{60}(p,n)Cu^{60}$	-6.6 ± 0.4	threshold		Co 54a	-6.6 ± 0.4
$Ni^{60}(d,p)Ni^{61}$	(5.55 ± 0.1)	range ph pl		Pr 54	
$Ni^{61}(p,\gamma)Cu^{62}$	6.03 ± 0.06	scint spec		Go 57a	preliminary
$Ni^{62}(p,\gamma)Cu^{63}$	6.13 ± 0.03	scint spec		Go 57a	
$Ni^{64}(p,\gamma)Cu^{65}$	7.42 ± 0.03	scint spec		Go 57a	
$Cu^{63}(\gamma,n)Cu^{62}$	-10.78 ± 0.05	threshold	D^2, F^{19}, $O^{16}(\gamma,n)$thresholds	Be 56	-10.65 ± 0.06
	-10.73 ± 0.05	threshold	N^{14}, $F^{19}(\gamma,n)$thresholds	Ro 55	
	-10.54 ± 0.04	threshold	C^{12},N^{14}, $O^{16}(\gamma,n)$thresholds	De 55	
$Cu^{63}(\gamma,2n)Cu^{61}$	-20.0 ± 0.5	threshold		Be 54	
$Cu^{63}(p,\gamma)Zn^{64}$	7.69 ± 0.04	scint spec		Go 57	preliminary
$Cu^{63}(p,n)Zn^{63}$	-4.147 ± 0.008	threshold	$Li^7(p,n)Be^7$	Ki 55	-4.149 ± 0.004
	-4.149 ± 0.004	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Br 55b	
$Cu^{65}(\gamma,n)Cu^{64}$	-9.94 ± 0.08	threshold	D^2, O^{16}, $F^{19}(\gamma,n)$thresholds	Be 56	-9.95 ± 0.08
$Cu^{65}(n,\gamma)Cu^{66}$	(7.01 ± 0.02)	pr spec	absolute	Ba 53, Qu 56a	(7.01 ± 0.02) tgt isotope uncertain
$Cu^{65}(p,\gamma)Zn^{66}$	8.85 ± 0.04	scint spec		Go 57	
$Cu^{65}(p,n)Zn^{65}$	-2.137 ± 0.005	threshold	$Li^7(p,n)Be^7$	Ki 55	$-2.132_7\pm0.001_3$
	-2.136 ± 0.004	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Br 55b	
	-2.131 ± 0.005	resonance	$Li^7(p,n)Be^7$	Ma 56	
	$-2.132_2\pm0.001_5$	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Ma 56b	
$Zn^{64}(\gamma,2n)Zn^{62}$	-20.35 ± 0.35	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$thresholds	De 55	
$Zn^{64}(\gamma,n)Zn^{63}$	-11.6	threshold		Ka 51	-11.60 ± 0.06
	-11.58 ± 0.06	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$thresholds	De 55	
$Zn^{64}(n,\gamma,\gamma)Zn^{65}$	(7.928 ± 0.007)			Ki 53, Wa 55	tgt isotope
	(7.990 ± 0.008)			Ki 53, Qu 56a	uncertain
$Zn^{66}(n,\gamma,\gamma)Zn^{67}$	(7.03 ± 0.02)			Ki 53, Qu 56a	tgt isotope uncertain
$Zn^{67}(n,\gamma,\gamma)Zn^{68}$	(10.22 ± 0.01)			Wa 55, Qu 56a	tgt isotope uncertain
$Zn^{67}(p,n)Ga^{67}$	-1.777 ± 0.005^{b}	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Ch 57	-1.781 ± 0.003
$Zn^{68}(n,\gamma)Zn^{69}$	(6.49 ± 0.02)			Ki 53, Qu 56a	tgt isotope uncertain
$Zn^{68}(p,n)Ga^{68}$	-3.694 ± 0.006	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Br 55b	-3.700 ± 0.005
	-3.704 ± 0.005^{b}	threshold	$Li^7(p,n)Be^7$, $E_T=1.8811$	Ch 57	
$Ga^{69}(n,\gamma)Ga^{70}$	7.733 ± 0.020	pr spec	absolute	Ba 56a	
$Ge^{70}(\gamma,n)Ge^{69}$	-12.1 ± 0.2	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$thresholds	De 55	
$Ge^{76}(\gamma,n)Ge^{75}$	-9.3	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$thresholds	De 55	
$Kr^{78}(d,p)Kr^{79}$	5.98 ± 0.05	range ph pl	$N^{14}(d,p)N^{15}$	Bl 56	
$Kr^{80}(d,p)Kr^{81}$	5.63 ± 0.10	range ph pl	$N^{14}(d,p)N^{15}$	Bl 56	
$Rb^{85}(\gamma,n)Rb^{84}$	-9.26 ± 0.15	threshold	O^{16}, $F^{19}(\gamma,n)$thresholds	To 56	
$Rb^{87}(\gamma,n)Rb^{86}$	-10.14 ± 0.15	threshold	O^{16}, $F^{19}(\gamma,n)$thresholds	To 56	
$Sr^{86}(\gamma,n)Sr^{85}$	-11.5	threshold		Ye 55	
$Sr^{86}(n,\gamma,\gamma)Sr^{87}$	8.417 ± 0.018			Wa 55	$8.417\pm0.018^{+}$
$Sr^{87}(n,\gamma,\gamma)Sr^{88}$	11.14 ± 0.05			Wa 55	$11.14\pm0.05^{+}$
$Y^{89}(d,p)Y^{90}$	4.41 ± 0.05	scint spec		Wa 54	
$Zr^{90}(\gamma,n)Zr^{89}$	-11.78 ± 0.09		$Cu^{63}(\gamma,n)Cu^{62}$, $E_T=10.73$	Ax 56	-11.78 ± 0.09
	The Q values for this reaction in reference (1) are now attributed to excitation of the 0.588 level in Zr^{89} (Ax 56).				

TABLE I.—*Continued.*

Reaction	Measured Q value (Mev)	Method	Calibration energy standard	Reference	Average Q value (Mev)
$Nb^{93}(p,n)Mo^{93}$	-1.27 ± 0.04	ph pl		Pa 54	-1.27 ± 0.04
	Value quoted in reference 1 is not ground-state transition.				
$Mo^{92}(\gamma,n)Mo^{91}$	(-13.5 ± 0.4)	threshold		Ba 45	-13.1 ± 0.1
	(-13.28 ± 0.15)	threshold		Ha 49a	
	-13.1 ± 0.1	threshold		Ka 53, Ka 53a	
$Mo^{92}(n,2n)Mo^{91}$	-12.34	threshold		Br 53	
$Rh^{103}(p,n)Pd^{103}$	(-1.53)	ph pl		Pa 54	
$Ag^{107}(\gamma,n)Ag^{106}$	-9.45 ± 0.05	threshold	D^2, F^{19}, $O^{16}(\gamma,n)$ thresholds	Be 56	-9.50 ± 0.06
	-9.57 ± 0.06	threshold	C^{12}, O^{16}, Cu^{63}, $Ag^{109}(\gamma,n)$-thresholds	Ba 55	
$Ag^{109}(\gamma,n)Ag^{108}$	-9.17 ± 0.06	threshold	D^2, F^{19}, $O^{16}(\gamma,n)$thresholds	Be 56	-8.90 ± 0.18
	-8.78 ± 0.04	threshold		De 55	
$Cd^{111}(p,n)In^{111}$	-2.37 ± 0.20	threshold		Mc 51c	
$Cd^{113}(n,\gamma)Cd^{114}$	9.04 ± 0.03	compt spec		Ad 55	9.046 ± 0.008
$Ba^{138}(d,p)Ba^{139}$	2.493 ± 0.010			Pa 55	2.493 ± 0.010
$Ba^{138}(n,\gamma)Ba^{139}$	(4.70 ± 0.03)	pr spec	absolute	Ki 53c, Pa 55	tgt isotope uncertain
$Pr^{141}(d,p)Pr^{142}$	3.42 ± 0.30	scint spec		Wa 54	
$Sm^{144}(\gamma,n)Sm^{143}$	-9.60 ± 0.05	threshold	C^{12}, N^{14}, $O^{16}(\gamma,n)$thresholds	De 55	
	-9.6	threshold		Si 56	
$Sm^{149}(n,\gamma,\gamma)Sm^{150}$	8.00 ± 0.03			Ad 55	8.00 ± 0.03
$Gd^{155}(n,\gamma)Gd^{156}$	(7.78 ± 0.05)	pr spec	absolute	Ki 53c	tgt isotope uncertain
$Gd^{157}(n,\gamma)Gd^{158}$	(7.36 ± 0.05)	pr spec	absolute	Kr 53c	tgt isotope uncertain
$Hf^{177}(\gamma,n)Hf^{176}$	-6.70 ± 0.09	threshold	O^{16}, $F^{19}(\gamma,n)$thresholds	To 56	
$Hf^{179}(\gamma,n)Hf^{178}$	-6.52 ± 0.12	threshold	O^{16}, $F^{19}(\gamma,n)$thresholds	To 56	
$Pt^{194}(n,\gamma)Pt^{195}$	(6.07 ± 0.04)	pr spec	absolute	Ki 53c	tgt isotope uncertain
$Hg^{199}(n,\gamma)Hg^{200}$	8.03 ± 0.03			Ad 55	8.03 ± 0.03
$Bi^{209}(p,2n)Po^{208}$	-9.65 ± 0.08	threshold	range Al, Sm 47	An 56	

[a] This Q value has been calculated specifically for this compilation from the experimental data, using accurate masses.
[b] This Q value has been corrected for the $Li^7(p,n)Be^7$ threshold energy of 1.8811 Mev.
[c] This Q value has been omitted from the weighted average.
* This average contains a later correction to the value originally reported.
\+ This average value was calculated by including the measured Q value for the inverse reaction.

List of abbreviations used for experimental methods.

ang corr	angular correlation	ph pl	range in photographic plates
compt spec	Compton electron spectrometer	pr spec	pair spectrometer
el spec	electrostatic spectrometer	pulse ht	pulse height
E_T	threshold energy	range Al	range in aluminum
mag spec	magnetic spectrometer	scint spec	scintillation spectrometer

given. When a correction to the original value has been reported by another author, a second reference with an asterisk has been added. For some measurements, such as a (n,γ,γ) reaction, two references are given. The first reference contains the measurements of the neutron capture gamma rays, while the second contains the assignment of the gamma rays to transitions for specific isotopes, and, also, other determinations of level energies if necessary for the calculation of the energy of the ground-state transition (e.g., see references 6 and 10).

In Column 6 are listed the weighted average values, calculated in the same manner as before. Measured values listed in the original compilation are included in the calculations of the weighted average values given in this supplement. Following the bibliography, a few additional corrections to the original compilation are listed, if not already included in this supplement. No attempt has been made to list revised references for those given previously if no alteration in the Q values reported were made in the later reference. However, all references listed in Table I of this supplement are included in the bibliography for the sake of completeness.

BIBLIOGRAPHY

Ad 55 Adyasevich, Groshev, and Demidov, Conf. Acad. Sci. U.S.S.R. *Peaceful Uses of Atomic Energy, Moscow, 1955*, Session Div. Phys. Math. Sci. English translation U.S.A.E.C. (1956), p. 195.

Ad 56 Adyasevich, Groshev, Demidov, and Lutsenko, Soviet J. Atomic Energy 1, 171 (1956); J. Nuclear Energy II, 3, 325 (1956).

Ad 56a Adyasevich, Groshev, and Demidov, Soviet J. of Atomic Energy 1, 183 (1956); J. Nuclear Energy II, 3, 258 (1956).

Ag 56 Ager-Hanssen, Lonsjo, and Nordhagen, Phys.Rev. **101**, 1779 (1956).

Ag 54a K. Ahnlund, Arkiv Fysik **7**, 459 (1954).

Ah 54b K. Ahnlund and C. Mileikowsky, Arkiv Fysik **8**, 161 (1954).

Ah 54c K. Ahnlund, Phys. Rev. **96**, 999 (1954).

Ah 54d Ahnlund, Thulin, and Pauli, Arkiv Fysik **8**, 489 (1954).

Aj 55 Ajzenberg, Rubin, and Likely, Phys. Rev. **99**, 654(A) (1955) and verbal report.

Aj 56 Ajzenberg-Selove, Johnson, Rubin, and Mazari, Phys. Rev. **103**, 356 (1956).

Aj 57 Ajzenberg-Selove, Bullock, and Almqvist, Bull. Am. Phys. Soc. Ser. II, **2**, 28 (1956).

Al 54 Allen, Almqvist, Dewan, and Pepper, Phys. Rev. **96**, 684 (1954).

Al 55 Almqvist, Clarke, and Paul, Phys. Rev. **100**, 1265(A) (1955).

An 56 Andre, Huizenga, Mech, Ramler, Rauh, and Rocklin, Phys. Rev. **101**, 645 (1956).

Ax 56 P. Axel and J. D. Fox, Phys. Rev. **102**, 400 (1956).

Ba 45 G. C. Baldwin and H. W. Koch, Phys. Rev. **67**, 1 (1945).

Ba 53 G. A. Bartholomew and B. B. Kinsey, Phys. Rev. **89**, 386 (1953).

Ba 53c S. J. Bame, Jr., Phys. Rev. **92**, 1096 (1953).

Ba 55 R. Basile and C. Schuhl, J. phys. radium **16**, 372 (1955).

Ba 55a R. Basile and C. Schuhl, Compt. rend. **240**, 2399 (1955).

Ba 56 G. A. Bartholomew and P. J. Campion, Bull. Am. Phys. Soc. Ser. II, **1**, 324 (1956).

Ba 56a G. A. Bartholomew (private communication).

Be 49 H. A. Bethe, AECU-346 (1949).

Be 50d H. Bethe, Revs. Modern Phys. **22**, 213 (1950).

Be 52 R. E. Beneson, Phys. Rev. **90**, 420 (1952).

Be 54 A. I. Berman and K. L. Brown, Phys. Rev. **96**, 83 (1954).

Be 55 E. H. Bellamy and F. C. Flack, Phil. Mag. **46**, 341 (1955).

Be 56 Bendel, McElhinney, and Tobin, Bull. Am. Phys. Soc. Ser. II, **1**, 192 (1956).

Bi 54a H. Bichsel and R. F. Mosley, Phys. Rev. **94**, 764(A) (1954).

Bi 55 J. R. Bird, Australian J. Phys. **8**, 314 (1955).

Bi 55a Bigham, Allen, and Almqvist, Phys. Rev. **99**, 631(A) (1955).

Bl 56 N. C. Blais and W. W. Watson, Phys. Rev. **104**, 202 (1956).

Bo 51 T. W. Bonner and J. W. Butler, Phys. Rev. **83**, 1091 (1951).

Bo 56 Bockelman, Leveque, and Buechner, Phys. Rev. **104**, 456 (1956).

Br 53 J. E. Brolley, Phys. Rev. **89**, 877 (1953).

Br 54 C. M. Braams, Phys. Rev. **94**, 763(A) (1954).

Br 54b C. M. Braams, Phys. Rev. **95**, 650 (1954).

Br 54c C. P. Browne, Phys. Rev. **95**, 860(L) (1954).

Br 55 C. P. Browne and W. C. Cobb, Phys. Rev. **99**, 644(A) (1955).

Br 55a C. P. Browne, Mass. Inst. Technol. Annual Progr. Rept. p. 117 (1955).

Br 55b Brugger, Bonner, and Marion, Phys. Rev. **100**, 84 (1955).

Br 56 C. M. Braams, Ph.D. thesis, University of Utrecht (1956).

Br 56a Bromley, Almqvist, Gove, Litherland, Paul, and Ferguson, PD-264, Atomic Energy of Canada, Ltd. (October, 1956).

Br 56b C. M. Braams, Phys. Rev. **103**, 1310 (1956).

Bu 53 Buechner, Sperduto, Browne, and Bockelman, Phys. Rev. **91**, 1502 (1953).

Bu 55 Buechner, Braams, and Sperduto, Phys. Rev. **100**, 1387 (1955).

Bu 56 Buechner, Mazari, and Sperduto, Phys. Rev. **101**, 188 (1956).

Bu 56a J. W. Butler, Bull. Am. Phys. Soc. Ser. II, **1**, 94 (1956) and private communication.

Bu 56b B. S. Burton and R. M. Williamson, Bull. Am. Phys. Soc. Ser. II, **1**, 264 (1956) and private communication from R. M. Williamson.

Bu 56c W. W. Buechner and A. Sperduto, Mass. Inst. Technol. Annual Progr. Rept., p. 107 (1956).

Bu 56d Butler, Dunning, and Bondelid, Bull. Am. Phys. Soc. Ser. II, **1**, 327 (1956).

Ca 56 P. J. Campion and G. A. Bartholomew, Bull. Am. Phys. Soc. Ser. II, **1**, 28 (1956).

Ch 56 Chadwick, Alexander, and Warren, Can. J. Phys. **34**, 381 (1956).

Ch 57 R. A. Chapman and J. C. Slattery, Phys. Rev. **105**, 633 (1957).

Co 54a Cohen, Newman, Charpie, and Handley, Phys. Rev. **94**, 620 (1954).

Co 55 Cook, Marion, and Bonner [private communication to F. Ajzenberg and T. Lauritsen, Revs. Modern Phys. **27**, 106 (1955)].

Co 56 D. R. Connors and W. C. Miller, Bull. Am. Phys. Soc. Ser. II, **1**, 340 (1956).

Cr 56 D. S. Craig, Phys. Rev. **101**, 1479 (1956).

Cr 56a Craig, Cross, and Jarvis, Phys. Rev. **103**, 1427 (1956).

De 55 de Sousa Santos, Goldemberg, Pieroni, Silva, Borello, Villaca, and Lopes, *Proceedings of International Conference on Peaceful Uses of Atomic Energy, Geneva, 1955* (United Nations, New York, 1956), Vol. II, p. 169.

Do 56 W. T. Doyle and A. B. Robbins, Phys. Rev. **101**, 1056 (1956).

Do 56a Douglas, Broer, Chiba, Herring, and Silverstein, Phys. Rev. **104**, 1059 (1956).

Du 56 Dunning, Butler, and Bondelid, Bull. Am. Phys. Soc. Ser. II, **1**, 328 (1956).

El 51c F. A. El-Bedewi, Proc. Phys. Soc. (London) **A64**, 1079 (1951).

El 52 El-Bedewi, Middleton, and Tai, Nature **169**, 235 (1952).

El 54 R. B. Elliot and D. J. Livesay, Proc. Roy. Soc. (London) **A224**, 129 (1954).

El 55 F. A. El-Bedewi and M. A. El Wahab, Proc. Phys. Soc. (London) **A68**, 754 (1955).

El 56 F. A. El-Bedewi, Proc. Phys. Soc. (London) **A69**, 221 (1956).

El 56a A. J. Elwyn and F. B. Shull, Bull. Am. Phys. Soc. Ser. II, **1**, 281 (1956).

El 57 F. A. El-Bedewi and I. Hussein, Proc. Phys. Soc. (London) **A70**, 233 (1957).

En 54 Endt, Kluyver, and Van der Leun, Physica **20**, 1299 (1954).

En 54a Endt, Kluyver, and Van der Leun, Phys. Rev. **95**, 580(L) (1954).

En 56 Endt, Paris, Sperduto, and Buechner, Phys. Rev. **103**, 961 (1956).

En 57 P. M. Endt and C. H. Paris, Phys. Rev. **106**, 764 (1957).

Fa 55 W. J. Fader, Mass. Inst. Technol. Annual Progr. Rept., p. 123 (1955).

Fl 53 F. C. Flack and J. B. Warren, Proc. Roy. Soc. (Canada) **47**, 131(A) (1953).

Fo 54 G. M. Foglesong and D. G. Foxwell, Phys. Rev. **96**, 1001 (1954).

Fo 56 Fowler, Lauritsen, Lauritsen, and Cook (private communication).

Fr 54a Freemantle, Grotdal, Gibson, McKeague, Prowse, and Rotblat, Phil. Mag. **45**, 1090 (1954); erratum **45** (November, 1954).

Gi 54 Gibson, Prowse, and Rotblat, Nature **173**, 1180 (1954).

Gi 55 Gibbons, Macklin, and Schmitt, Phys. Rev. **100**, 167 (1955).

Go 57 C. R. Gossett, Bull. Am. Phys. Soc. Ser. II, **2**, 69 (1957).

Go 57a C. R. Gossett and J. W. Butler (private communication).

Gr 54 G. W. Greenlees, Proc. Phys. Soc. (London) **67A**, 1107 (1954).

Gr 55 Green, Scanlon, and Willmott, Proc. Phys. Soc. (London) **68A**, 386 (1955).

Gr 55a Green, Singh, and Willmott, Phil. Mag. **46**, 982 (1955).

Gr 55b Groshev, Adyasevich, and Demidov, *Proceedings of International Conference on Peaceful Uses of Atomic Energy, Geneva, 1955* (United Nations, New York, 1956), Vol. II, p. 39.

Gr 55c G. M. Griffiths and T. B. Warren, Proc. Phys. Soc. (London) **68A**, 781 (1955).

Gr 57 H. G. Graetzer and A. B. Robbins, Phys. Rev. **105**, 1570 (1957).

Ha 49a Hanson, Duffield, Knight, Diven, and Palevsky, Phys. Rev. **76**, 578 (1949).

Ha 54 Haslam, Roberts, and Robb, Can. J. Phys. **32**, 361 (1954).

Ha 54a Halpern, Nathan, and Yergin, Phys. Rev. **95**, 1529 (1954).

Ha 56 Hassler, Steigert, and Pieper, Bull. Am. Phys. Soc. Ser. II, **1**, 280 (1956).

He 55 N. P. Heydenburg and G. M. Temmer, Phys. Rev. **99**, 617(A) (1955).

Hj 53 E. Hjalmar and H. Slatis, Arkiv Fysik **6**, 451 (1953).

Hj 53a E. Hjalmar and H. Slatis, Phys. Rev. **89**, 1151 (1953).

Ho 55 Holmgren, Hanscome, and Willet, Phys. Rev. **98**, 241(A) (1955).

Ih 55 M. A. Ihsan, Proc. Phys. Soc. (London) **A68**, 393 (1955).

Is 50 R. Ishiwari and K. Yuasa, Mem. Coll. Sci. Univ. Kyoto **A26**, 151 (1950).

Ja 55 James, Kubelka, Heiberg, and Warren, Can. J. Phys. **33**, 219 (1955).

Ja 56 N. Jarmie, Phys. Rev. **104**, 1683 (1956).

Jo 54 Jones, McEllistrem, Douglas, and Richards, Phys. Rev. **94**, 947 (1954).

Jo 56 C. H. Johnson [private communication to R. A. Chapman and J. C. Slattery, Phys. Rev. **105**, 633 (1957)].

Ju 54 J. J. Jung and C. K. Bockelman, Phys. Rev. **96**, 1353 (1954).

Ka 51 Katz, Johns, Baker, Haslam, and Douglas, Phys. Rev. **82**, 271 (1951).

Ka 53 Katz, Baker, and Montalbatti, Can. J. Phys. **31**, 250 (1953).

Ka 53a L. Katz [private communication to J. E. Brolley, Phys. Rev. **89**, 877 (1953)].

Ka 54 Katz, Haslam, Goldemberg, and Taylor, Can. J. Phys. **32**, 580 (1954).

Ka 55 Kavanagh, Mills, and Sherr, Phys. Rev. **97**, 248(L) (1955).

Kh 53 L. M. Khromchenko, Doklady Akad. Nauk. SSSR **93**, 451 (1953).

Kh 54 L. M. Khromchenko, Doklady Akad. Nauk. SSSR **94**, 1037 (1954).

Kh 55 L. M. Khromchenko and V. A. Blinov, Soviet Physics JETP **1**, 596 (1955).

Ki 53 B. B. Kinsey and G. A. Bartholomew, Phys. Rev. **89**, 375 (1953).

Ki 53c B. B. Kinsey and G. A. Barthollmew, Can. J. Phys. **31**, 1051 (1953).

Ki 53d Kington, Bair, Carlson, and Willard, Phys. Rev. **89**, 530 (1953).

Ki 55 Kington, Bair, Cohn, and Willard, Phys. Rev. **99**, 1393 (1955).

Kl 54 Kluyver, Van der Leun, and Endt, Phys. Rev. **94**, 1795(L) (1955); Physica **20**, 1287 (1955).

La 47 Lattes, Fowler, and Cuer, Proc. Phys. Soc. (London) **A59**, 883 (1947).

Le 55 Levine, Bender, and McGruer, Phys. Rev. **97**, 1249 (1955).

Le 56 L. L. Lee, Jr., and F. P. Mooring, Phys. Rev. **104**, 1342 (1956).

Li 37 M. S. Livingston and H. A. Bethe, Revs. Modern Phys. **9**, 245 (1937).

Li 55 J. G. Likely, Phys. Rev. **98**, 1538(A) (1955).

Li 57 Lidofsky, Weil, and Jones, Bull. Am. Phys. Soc. Ser. II, **2**, 182 (1957).

Lo 52 Lovington, McCue, and Preston, Phys. Rev. **85**, 585 (1952).

Ma 55a Marion, Brugger, and Bonner, Phys. Rev. **100**, 46 (1955).

Ma 55b Marion, Bonner, and Cook, Phys. Rev. **100**, 91 (1955).

Ma 56 J. M. Marion and R. A. Chapman, Phys. Rev. **101**, 283 (1956).

Ma 56a H. Mark and C. Goodman, Phys. Rev. **101**, 768 (1956).

Ma 56b J. B. Marion and R. W. Kavanagh, Phys. Rev. **104**, 107 (1956).

Mc 51c C. L. McGinnis, Phys. Rev. **81**, 734 (1951).

Mc 55 McFarland, Shull, Elwyn, and Zeidman, Phys. Rev. **99**, 655(A) (1955).

Mi 54 C. Mileikowsky, Arkiv Fysik **7**, 89 (1954).

Mi 54a C. Mileikowsky, Arkiv Fysik **7**, 117 (1954).

Mi 54b C. Mileikowsky and K. Ahnlund, Phys. Rev. **96**, 996 (1954).

Mo 54 H. T. Motz, Phys. Rev. **93**, 925(A) (1954), probable error is our estimate.

No 56 Nonaka, Morita, Kawai, Ishimatsu, Suematsu, Takeshita, Nakajima, and Wakuda, J. Phys. Soc. Japan **11**, 1 (1956).

No 57 E. Norbeck, Jr., Phys. Rev. **105**, 204 (1957).

Og 50a Ogle, Brown, and Carson, Phys. Rev. **78**, 63 (1950).

Pa 54 R. Patterson, Phys. Rev. **95**, 303(A) (1954).

Pa 54a Pauli, Ahnlund, and Mileikowsky, Arkiv Fysik **8**, 213 (1954).

Pa 55 Paris, Buechner, and Endt, Phys. Rev. **100**, 1317 (1955).

Pa 55a R. T. Pauli, Arkiv Fysik **9**, 571 (1955).

Pa 56 A. Papkow, Z. Naturforsch. **11**, 776 (1956).

Pe 52 Pepper, Allen, Almqvist, and Dewan, Phys. Rev. **85**, 155 (1952).

Pi 52a G. F. Pieper, Phys. Rev. **88**, 1299 (1952).

Pi 55 Pieper, Stanford, and von Herrmann, Phys. Rev. **98**, 1185(A) (1955).

Pi 56 G. F. Pieper and G. S. Stanford, Phys. Rev. **101**, 672 (1956).

Pl 55 H. S. Plendl and F. E. Steigert, Phys. Rev. **98**, 1538(A) (1955).

Pr 52 Pruitt, Hanna, and Swartz, Phys. Rev. **87**, 534 (1952).

Pr 53 Pruitt, Swartz, and Hanna, Phys. Rev. **92**, 1456 (1953).

Pr 54 W. W. Pratt, Phys. Rev. **95**, 1517 (1954).

Qu 56 A. R. Quinton and W. T. Doyle, Phys. Rev. **101**, 669 (1956).

Qu 56a Quisenberry, Scolman, and Nier, Phys. Rev. **104**, 461 (1956).

Re 54 G. C. Reid, Proc. Phys. Soc. (London) **A67**, 466 (1954).

Re 55 J. B. Reynolds, Phys. Rev. **98**, 1289 (1955).

Ri 51 Richards, Johnson, Ajzenberg, and Laubenstein, Phys. Rev. **83**, 994 (1951).

Ro 51c J. Rotblat, Nature **167**, 550 (1951).

Ro 55 Robinson, McPherson, Greenberg, Katz, and Haslam (unpublished), quoted in A. S. Penfold and B. M. Spicer, Phys. Rev. **100**, 1377 (1955).

Ro 56 A. B. Robbins, Phys. Rev. **101**, 1373 (1956).

Ru 56 Rubin, Johnson, and Reynolds, Phys. Rev. **104**, 1444 (1956).

Sa 56 R. W. Sanders, Phys. Rev. **104**, 1434 (1956).

Sc 54 Schnelder, Martin, Sempert, and Sutter, Helv. Phys. Acta **27**, 127A (1954).

Sc 55 J. P. Schiffer, Phys. Rev. **97**, 428 (1955).

Sc 56 M. J. Scott and R. E. Segel, Phys. Rev. **102**, 1557 (1956).

Sc 56a Schwartz, Corbett, and Watson, Phys. Rev. **101**, 1370 (1956).

Sh 51 Sher, Halpern, and Mann, Phys. Rev. **84**, 387 (1951).

Sh 51a Sher, Halpern, and Stephens, Phys. Rev. **81**, 154 (1951).

Si 56 E. Silva and J. Goldemberg, Nuovo cimento, **3**, 12 (1956).

Sm 47 J. H. Smith, Phys. Rev. **71**, 32 (1947).
Sp 54 Sperduto, Buechner, Bockelman, and Browne, Phys. Rev. **96**, 1316 (1954).
Sp 55 A. Sperduto and W. W. Buechner, Phys. Rev. **100**, 961(A) (1955); Mass. Inst. Technol. Progr. Rept., p. 56 (November, 1955).
Sp 56 A. Sperduto and W. W. Buechner, Bull. Am. Phys. Soc. Ser. II, **1**, 223 (1956).
Sp 56a A. Sperduto and W. W. Buechner, Mass. Inst. Technol. Annual Progr. Rept., p. 111 (1956).
Sq 56 Squires, Bockelman, and Bueckner, Phys. Rev. **104**, 413 (1956).
St 52 P. H. Stelson and W. M. Preston, Phys. Rev. **86**, 807 (1952).
St 56 G. S. Stanford and G. F. Pieper, Phys. Rev. **103**, 637 (1956).
Su 55 S. Subotic and B. Maglic, Phil. Mag. **46**, 805(L) (1955).
Th 54 Thirion, Cohen, and Whaling, Phys. Rev. **96**, 850(A) (1954).
To 56 R. Tobin and J. McElhinney, Bull. Am. Phys. Soc. Ser. II, **1**, 340 (1956).
Tr 55 P. B. Treacy, Proc. Phys. Soc. (London) **68A**, 204 (1955).
Va 56 Van Patter, Swann, Porter, and Mandeville, Phys. Rev. **103**, 656 (1956).
Va 56a C. Van der Leun and P. M. Endt, Physica **22**, 1234 (1956).
Va 57 Van Patter, Porter, and Rothman, Phys. Rev. **106**, 1016 (1957).
Wa 54 N. S. Wall, Phys. Rev. **96**, 664 (1954).
Wa 55 Way, King, McGinnis, and van Lieshout, "Nuclear Level Schemes," U.S. Atomic Energy Commission TID-5300 (1955).
Wa 56 Warren, Alexander, and Chadwick, Phys. Rev. **101**, 242 (1956).
Wh 55 W. Whaling [private communication to R. T. Pauli, Arkiv Fysik **9**, 571 (1955)].
Wi 52 Willard, Bair, Kington, Hahn, and Green, Phys. Rev. **85**, 849 (1952).
Wi 52a Willard, Kington, and Bair, Phys. Rev. **86**, 259 (1952).
Ye 55 P. F. Yergin and B. P. Fabricand, Phys. Rev. **100**, 1269(A) (1955).

Corrections to "Nuclear Disintegration Energies" [Revs. Modern Phys. 26, 402 (1954)].

$Li^7(p,\alpha)He^4$ The average Q value should read 17.345 ± 0.006 Mev.

$Si^{28}(p,n)P^{28}$ The Q value measured by Br 54a should be -14.9 ± 0.6 Mev.
The average Q value should read -15.0 ± 0.4 Mev.

$Cr^{54}(p,n)Mn^{54}$ The Q value measured by Lo 52 should read -2.162 ± 0.005 Mev.

$Cu^{65}(n,\gamma)Cu^{66}$ Assignment of 7.634 Mev γ to this reaction is now thought to be incorrect (Qu 56a, Wa 55).

Be 50d H. A. Bethe, Revs. Modern Phys. **22**, 213 (1950).
El 51d F. A. El-Bedewi, Proc. Phys. Soc. (London) **A64**, 584 (1951).

Mass Spectroscopic Atomic Mass Differences. II.*†

HENRY E. DUCKWORTH

Department of Physics, Hamilton College, McMaster University, Hamilton, Ontario, Canada

I. INTRODUCTION

A BRIEF description is given of the mass spectroscopic methods in current use for the determination of atomic masses. The agreement between atomic masses so-derived and those obtained by studying the energy balance in nuclear reactions is also lightly touched upon, with particular reference to the mass of C^{12}. Finally, a table is given of the atomic mass differences which have been obtained mass spectroscopically during the past three years. This table supplements (and occasionally amends) a similar table which appeared[1] in the October, 1954, issue of this Journal.

II. PRECISION MASS SPECTROSCOPY

The mass spectroscopes employed in the study of atomic masses are necessarily high resolution instruments. With a single exception, those under construction or now in use are deflection instruments which take advantage of the double-focusing property which may be achieved by appropriately combining electrostatic and magnetic fields. The exception is the "mass synchrometer" of L. G. Smith[2] at the Brookhaven National Laboratory, in which measurements are made of the cyclotron frequencies of the motions of ions in a homogeneous magnetic field.

In a mass spectroscope, depending upon whether photographic or electrical detection is employed, "lines" or "peaks" are recorded. As a rule, in mass comparison work, *two* such lines or peaks are observed, representing two groups of ions whose specific charges are nearly equal. These constitute a "doublet," and it is the concern of the mass spectroscopist to ascertain, in terms of mass, the doublet spacing. Indeed, most of the data given in Table II, which is the principal *raison d'être* of this review, are simply mass differences for various doublets.

The precision with which such a mass difference can be determined is dependent upon the precision with which a line or peak may be located. Let us designate the mass width of a line or peak by Δm, a quantity which is directly proportional to the resolution $(\Delta m/m)$ of the mass spectroscope. In the case of photographic recording, the position of a mass spectral line may be determined to some fraction of its width, say 1/50 for an observer who is neither unduly optimistic nor unduly conservative. Thus, if a resolution of 1/20 000 be available, the mass of an atom may be determined with a precision of one part in a million. In practice the grain size ($\sim 10^{-4}$ cm) of the photographic plate sets a lower limit to the *actual* line width, with the result that, for a given size of mass spectroscope, the resolution cannot be improved beyond a certain point. This limit has been approximately reached by Mattauch and his collaborators[3] and by Ewald[4] who have achieved resolutions of 1/100 000 with mass spectrographs possessing dispersions of only 0.2 cm/1% mass difference. This corresponds to an actual line width of 2×10^{-4} cm.

With electrical recording, it has been demonstrated that it is possible to locate a peak with a precision of approximately 1/500 of its width, a tenfold improvement over the photographic case. This is done by the "peak-matching" technique introduced by Smith[5] and also employed, in a modified form, by Nier and his collaborators.[6] In this technique, by taking advantage of rapidly-responding detector systems, the two doublet peaks are made to appear on an oscilloscope screen on alternate sweeps. These two peaks are then brought into coincidence by adjustment of some circuit parameter (in Smith's case, a frequency; in Nier's case, a resistance) whose value gives the doublet mass difference. The peaks are thus "matched" by the human eye, an organ which can discern lack of coincidence with exceptionally keen discrimination. This peak-matching scheme is perhaps the most important single advance in precision mass spectroscopy since the discovery of the double-focusing principle in the mid-1930's. By this device, stated precisions have been regularly achieved of one part in 2×10^7 and, occasionally, of even one part in 10^8.

Partly because of the grain-size limitation in the case of photographic detection instruments but, even more, because of the difficulty of making and aligning the diminutive slits needed in small instruments if high resolution is to be attained, the current trend in precision mass spectroscopes is toward the bigger-and-better variety. Such instruments are larger than their predecessors by roughly an order of magnitude and are

* This work was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command and by the National Research Council of Canada.

† Prepared at the suggestion of the Subcommittee on Nuclear Constants of the Committee on Nuclear Science of the National Research Council.

[1] Duckworth, Hogg, and Pennington, Revs. Modern Phys. **26**, 463 (1954).

[2] L. G. Smith and C. C. Damm, Rev. Sci. Instr. **27**, 638 (1956).

[3] F. Everling, *Proceedings of the Mainz Conference on Atomic Masses* (Pergamon Press, New York, 1957).

[4] H. Ewald, Natl. Bur. Standards Circ. 522, 37 (1953).

[5] L. G. Smith and C. C. Damm, Phys. Rev. **90**, 324 (1953).

[6] Quisenberry, Scolman, and Nier, Phys. Rev. **102**, 1071 (1956).

TABLE I. Recent values for the mass of C^{12}.

Method	Investigator(s)	C^{12}	Error	Reference
Mass spectroscopic	Collins, Nier, and Johnson (1952)	12.003842	4	a
	Ogata and Matsuda (1953)	12.003844	6	b
	Mattauch and Bieri (1954)	12.0038231	33	Ma 54
	Smith (1955)	12.0038212	38	c
	Quisenberry, Scolman, and Nier (1955)	12.0038174	18	d
	Liebl and Ewald (1956)	12.003819	2	Li 56
	Kettner (1956)	12.003814	6	Ket 56
	Quisenberry, Scolman, and Nier (1956)	12.0038167	8	Qu 56
	Demirkhanov *et al.* (1956)	12.003820	5	De 56
	Smith (1957)	12.00381458	11	Sm 57
	Quisenberry, Giese, and Benson (1957)	12.0038156	4	Qu 57
Nuclear reactions	Li *et al.* (1951)	12.003804	17	e
	Wapstra (1955)	12.003803	5	f
	Mattauch *et al.* (1956)	12.0038000	39	g

[a] Collins, Nier, and Johnson, Phys. Rev. **86**, 408 (1952).
[b] K. Ogata and H. Matsuda, Phys. Rev. **89**, 27 (1953).
[c] L. G. Smith, Third Annual Meeting, Am. Soc. Testing Materials Committee E-14 on Mass Spectrometry (1955).
[d] Quisenberry, Scolman, and Nier, Phys. Rev. **100**, 1245(A) (1955).
[e] Li, Whaling, Fowler, and Lauritsen, Phys. Rev. **83**, 512 (1951).
[f] A. H. Wapstra, Physica **21**, 367 (1955).
[g] Mattauch, Waldmann, Bieri, and Everling, Z. Naturforsch. **11a**, 525 (1956).

under construction at Osaka University,[7,8] Harvard University,[9] the Max Planck Institute for Chemistry[10] and McMaster University.[11]

Although doublets are the traditional objects of study in precision mass spectroscopy, Johnson and Nier[12] have shown that electrical recording makes possible the determination of mass differences that are *much* larger than doublet spacings. This, with photographic detection, would require a knowledge of the dispersion of the mass spectrograph over a considerable range, a knowledge which no mass spectroscopist to date has had the temerity to aver that he possesses. With electrical recording, however, the ion groups are brought in turn to the same collector, say, by altering a resistance that determines the voltage across the electrostatic analyzer, as in Nier's case. Here the ions have each traveled identical paths at the time of detection, and the mass change in moving from one peak to another is proportional to the corresponding resistance change. The accuracy of this dispersion law has been repeatedly verified by determining directly the H^1 mass as, for example, from the $C_5H_8-C_5H_7$ mass difference. These "mass unit" (MU) mass differences do not carry with them the same precision as a doublet difference but they have already proved[13] an important means of determining neutron binding energies among the heavy atoms.

[7] K. Ogata and H. Matsuda, Z. Naturforsch. **10a**, 843 (1955).
[8] K. Ogata and H. Matsuda, *Proceedings of the Mainz Conference on Atomic Masses* (Pergamon Press, New York, 1957).
[9] K. T. Bainbridge and T. L. Collins, *Proceedings of the Mainz Conference on Atomic Masses* (Pergamon Press, New York, 1957).
[10] Everling, Hintenberger, König, Mattauch, Müller-Warmuth, and Wende, *Proceedings of the Mainz Conference on Atomic Masses* (Pergamon Press, New York, 1957).
[11] Duckworth, Kerr, and Bainbridge, *Proceedings of the Mainz Conference on Atomic Masses* (Pergamon Press, New York, 1957).
[12] W. H. Johnson and A. O. Nier, Phys. Rev. **105**, 1014 (1957).
[13] W. H. Johnson and V. B. Bhanot, Bull. Am. Phys. Soc. Ser. II, 2, 224 (1957).

III. MASS SPECTROSCOPIC *VERSUS* NUCLEAR REACTION MASSES—C^{12}

Three years ago, at the time the first compilation was prepared, the lack of agreement between the mass of C^{12} as determined mass spectroscopically and as derived from reaction data was a source of concern. In the interval, the absolute value of this discrepancy has been greatly reduced, in fact, from $\sim 40\,\mu$ MU to $\sim 15\,\mu$ MU. The discrepancy which remains, however, is still a real one which, statistically-speaking, is more significant than before. The evidence is presented in Table I.

Discrepancies exist elsewhere as well, but these are not so sharply defined, as the nuclides in question have not been the subject of such intense investigation. Giese and Benson,[14] at the University of Minnesota, have recently concluded a mass spectroscopic study of atomic masses in the region $31 \leqslant A \leqslant 55$. These results, when compared with the masses derived from nuclear reactions, show a discrepancy whose magnitude appears to be linearly dependent upon the distance by which the nuclide is removed from O^{16}. Such a discrepancy could be explained by assuming small errors in the determination of nuclear reaction Q's. These errors would be cumulative and, consequently, the total error would increase with the number of reaction stages connecting the nuclide in question to O^{16}. On the other hand, it is hard to imagine a systematic error in the mass spectroscopic work that could lead to this steadily-increasing type of discrepancy. True, an error in the mass of C^{12} is potentially present, but this is much smaller than the inconsistencies here referred to. In each case the mass of the atom is found by studying some particular doublet, that is, by a one-stage operation. Furthermore, the doublets employed by Giese and Benson are such that they do not

[14] C. F. Giese and Benson, Bull. Am. Phys. Soc. Ser. II, 2, 223 (1957).

TABLE II. Mass spectroscopic atomic mass differences.[a]

Element Z	A	Doublet	ΔM in mMU	Error	Reference
1 H	1	H_2-D	1.5483	10	De 56
		CH_2-N	12.5803	4	Qu 56
		CH_4-O	36.388	4	De 56
			36.3931	9	Qu 56
		NH_2-O	23.8164	5	Qu 56
		CH_4-NH_2	12.5804	5	Qu 56
		NH_3-OH	23.833	8	Ket 56
			23.8159	6	Qu 56
		NH_3-NDH	1.5478	4	Qu 56
		$C^{13}H_4-OH$	31.943	11	Ket 56
			31.9253	7	Sco 56
		H_2O-DO	1.5476	5	Qu 56
		H_2O-O^{18}	11.4033	21	Qu 56
		$N^{15}H_3-H_2O$	13.019	5	Ket 56
			13.0234	4	Sco 56
		$H_3O-H_2O^{17}$	3.601	6	Ket 56
		H_3O-HO^{18}	11.405	8	Ket 56
		$D_2O-H_2O^{18}$	8.3102	4	Qu 56
		D_2O-HF^{19}	16.8944	5	Sco 56
		C_2H_4-CO	36.3934	8	Qu 56
			36.3960	5	Qu 57
		$C_2H_4-N_2$	25.1585	6	Sco 56
		CH_3OH-O_2	36.3958	12	Qu 57
		$\frac{1}{2}C_4H_2O-O^{17}O$	11.224	9	Ket 56
		$\frac{1}{2}C_4H_4O-H_2S$	25.3926	9	Qu 56
		$\frac{1}{2}C_4H_4O-H_2O_2$	7.6312	8	Qu 57
		$\frac{1}{2}C_4H_4O-O^{18}O$	19.040	13	Ket 56
			19.0367	8	Qu 56
		$C_3H_4-A^{40}$	68.9344	13	Qu 56
			68.9346	11	Qu 57
		$C_3H_8-CO_2$	72.7870	16	Qu 56
			72.7932	16	Qu 57
D 2		H_2-D	1.5483	10	De 56
		D_2-He^4	25.600	2	De 56
		$D_3-\frac{1}{2}C$	42.298	7	De 56
		NH_3-NDH	1.5478	4	Qu 56
		H_2O-DO	1.5476	5	Qu 56
		CD_4-D_2O	33.29453	11	Sm 57
		ND_3-D_2O	22.26447	10	Sm 57
		CD_4-ND_3	11.03006	10	Sm 57
		CD_4-Ne^{20}	63.97896	13	Sm 57
		ND_3-Ne^{20}	52.94890	12	Sm 57
		D_2O-Ne^{20}	30.6872	7	Sco 56
			30.68443	12	Sm 57
		$CD_4-\frac{1}{2}A^{40}$	75.22626	17	Sm 57
		$D_2O-\frac{1}{2}A^{40}$	41.9390	13	Qu 56
			41.93173	15	Sm 57
		$ND_3-\frac{1}{2}A^{40}$	64.19620	16	Sm 57
		$C_3D_8-C_4D_2$	84.62526	22	Sm 57
2 He	4	D_2-He^4	25.600	2	De 56
		$He_2^4-\frac{1}{2}O$	‡7.7543	44	Ma 54
5 B	10	$B^{10}H-B^{11}$	11.450	12	Li 56
		$B^{10}B^{11}H_2-Na^{23}$	48.130	10	Li 56
		$B^{10}F_2^{19}-S^{32}O$	42.7730	17	Sco 56
	11	$B^{10}H-B^{11}$	11.450	12	Li 56
		$B^{11}H-C$	17.139	7	Li 56
		$B^{10}B^{11}H_2-Na^{23}$	48.130	10	Li 56
		$C_4H_4O-B^{11}F_3^{19}$	21.7052	13	Sco 56
6 C	12	$D_3-\frac{1}{2}C$	42.298	7	De 56
		CH_2-N	12.5803	4	Qu 56
		CH_4-NH_2	12.5804	5	Qu 56
		CH_4-O	36.388	4	De 56
			36.3931	9	Qu 56
		CD_4-D_2O	33.29453	11	Sm 57
		CD_4-ND_3	11.03006	10	Sm 57
		CD_4-Ne^{20}	63.97896	13	Sm 57
		$CD_4-\frac{1}{2}A^{40}$	75.22626	17	Sm 57
6 C	12	C_2H_4-CO	36.3934	8	Qu 56
			36.3960	5	Qu 57
		$C_2H_4-N_2$	25.1585	6	Sco 56
		N_2-CO	11.2355	6	Sco 56
		CH_3OH-O_2	36.3958	12	Qu 57
		$\frac{1}{2}C_4H_2O-O^{17}O$	11.224	9	Ket 56
		$\frac{1}{2}C_4H_4O-H_2S$	25.3926	9	Qu 56
		$\frac{1}{2}C_4H_4O-H_2O_2$	7.6312	8	Qu 57
		$\frac{1}{2}C_4H_4O-O^{18}O$	19.040	13	Ket 56
			19.0367	8	Qu 56
		$C_3H_4-A^{40}$	68.9344	13	Qu 56
			68.9346	11	Qu 57
		$C_3H_8-CO_2$	72.7870	16	Qu 56
			72.7932	16	Qu 57
		N_2O-CO_2	11.244	10	Kr 57
		C_4-SO	33.0269	13	Qu 56
		$C_3D_8-C_4D_2$	84.62526	22	Sm 57
	13	$C^{13}H_4-HO$	31.943	11	Ket 56
			31.9253	7	Sco 56
7 N	14	CH_2-N	12.5803	4	Qu 56
		NH_2-O	23.8164	5	Qu 56
		NH_3-OH	23.833	8	Ket 56
			23.8159	6	Qu 56
		CD_4-ND_3	11.03006	10	Sm 57
		ND_3-D_2O	22.26447	10	Sm 57
		ND_3-Ne^{20}	52.94890	12	Sm 57
		$ND_3-\frac{1}{2}A^{40}$	64.19620	16	Sm 57
		$C_2H_4-N_2$	25.1585	6	Qu 56
		N_2-CO	11.2355	6	Sco 56
		N_2O-CO_2	11.244	10	Kr 57
	15	CH_3-N^{15}	23.3652	9	Sco 56
		$N^{15}H_3-H_2O$	13.019	5	Ket 56
			13.0234	4	Sco 56
8 O	17	$OH-O^{17}$	3.6077	5	Sco 56
		$H_3O-H_2O^{17}$	3.601	6	Ket 56
		$\frac{1}{2}C_4H_2O-O^{17}O$	11.224	9	Ket 56
	18	H_2O-O^{18}	11.4033	21	Qu 56
		H_3O-HO^{18}	11.405	8	Ket 56
		$D_2O-H_2O^{18}$	8.3102	4	Qu 56
		$H_2O^{18}-HF^{19}$	8.582	2	Ket 56
		$H_2O^{18}-Ne^{20}$	22.392	5	Ket 56
			22.3770	6	Sco 56
		$HDO^{18}-Ne^{21}$	27.2482	7	Sco 56
		$\frac{1}{2}COO^{18}-Na^{23}$	7.2592	20	Sco 56
		$\frac{1}{2}C_4H_4O-O^{18}O$	19.040	13	Ket 56
			19.0367	8	Qu 56
9 F	19	$HDO-F^{19}$	18.4380	14	Sco 56
		$H_2O^{18}-HF^{19}$	8.582	2	Ket 56
		D_2O-HF^{19}	16.8944	5	Sco 56
		$B^{10}F_2^{19}-SO$	42.7730	17	Sco 56
		$C_4H_4O-B^{11}F_3^{19}$	21.7052	13	Sco 56
		$C_5H_9-CF_3^{19}$	75.2462	20	Sco 56
		$C_6H_{13}-Si^{28}F_3^{19}$	129.625	4	Sco 56
		$C_6H_{14}-Si^{29}F_3^{19}$	137.889	7	Sco 56
		$C_4H_7O_2-Si^{30}F_3^{19}$	75.6590	36	Sco 56
10 Ne	20	CD_4-Ne^{20}	63.97896	13	Sm 57
		ND_3-Ne^{20}	52.94890	12	Sm 57
		D_2O-Ne^{20}	30.6872	7	Sco 56
			30.68443	12	Sm 57
		$H_2O^{18}-Ne^{20}$	22.392	5	Ket 56
			22.3770	6	Qu 56
		$Ne^{20}-\frac{1}{2}A^{40}$	11.24730	18	Sm 57
	21	$\frac{1}{2}C_2H_2O-Ne^{21}$	11.429	5	Ket 56
		$HDO^{18}-Ne^{21}$	27.2482	7	Sco 56
	22	$\frac{1}{2}CO_2-Ne^{22}$	3.521	12	Ket 56
			3.5307	6	Sco 56

[a] Special symbols appearing in this table are defined in the second last paragraph of Sec. IV.

TABLE II.—*Continued.*

Element Z		A	Doublet	ΔM in mMU	Error	Reference
11	Na	23	$B^{10}B^{11}H_2-Na^{23}$	48.130	10	Li 56
			$\frac{1}{2}COO^{18}-Na^{23}$	7.2592	20	Sco 56
12	Mg	24	C_2-Mg^{24}	14.9621	11	Sco 56
		25	C_2H-Mg^{25}	21.9944	10	Sco 56
		26	$C_2H_2-Mg^{26}$	33.0676	10	Sco 56
13	Al	27	$C_2H_3-Al^{27}$	41.9548	23	Sco 56
14	Si	28	$C_6H_{13}-Si^{28}F_3^{19}$	129.625	4	Sco 56
		29	$C_6H_{14}-Si^{29}F_3^{19}$	137.889	7	Sco 56
		30	$C_4H_7O_2-Si^{30}F_3^{19}$	75.6590	36	Sco 56
15	P	31	$O_2-P^{31}H$	8.2423	6	Ge 57
16	S	32	O_2-S	17.756	10	Sch 56
				17.7599	9	Qu 56
				17.7623	11	Qu 57
			$HS-S^{33}$	8.448	25	Sch 56
			$\frac{1}{2}C_4H_4O-H_2S$	25.3926	9	Qu 56
		33	$HS-S^{33}$	8.448	25	Sch 56
			$C_4H-S^{33}O$	41.4602	15	Ge 57
		34	H_2S-S^{34}	19.851	10	Sch 56
			$C_3-H_2S^{34}$	16.466	10	Sch 56
			$C_4H_2-S^{34}O$	52.9889	15	Ge 57
		36	$C_4H_4-S^{36}O$	69.3175	35	Ge 57
17	Cl	35	$HCl^{35}-A^{36}$	9.1346	9	Ge 57
			$C_5H_{10}-Cl_2^{35}$	140.5850	34	Ge 57
		37	$HCl^{37}-A^{38}$	11.0001	10	Ge 57
			$C_6H_2-Cl_2^{37}$	83.8692	23	Ge 57
18	A	36	$H_2O-\frac{1}{2}A^{36}$	(26.7937	6)	Qu 56
			C_3-A^{36}	(32.4729	20)	Qu 56
			$HCl^{35}-A^{36}$	9.1346	9	Ge 57
		38	$HCl^{37}-A^{38}$	11.0001	10	Ge 57
		40	$CD_4-\frac{1}{2}A^{40}$	75.22626	17	Sm 57
			$ND_3-\frac{1}{2}A^{40}$	64.19620	16	Sm 57
			$D_2O-\frac{1}{2}A^{40}$	41.9390	13	Qu 56
				41.93173	15	Sm 57
			$Ne^{20}-\frac{1}{2}A^{40}$	11.24730	18	Sm 57
			$C_3H_4-A^{40}$	68.9344	13	Qu 56
				68.9346]	11	Qu 57
19	K	39	$C_3H_3-K^{39}$	59.7621	20	Li 56
				59.7819	15	Ge 57
			C_2HN-K^{39}	‡47.58	8	He 51
		40	$C_3H_4-K^{40}$	67.3178	21	Ge 57
		41	$C_3H_5-K^{41}$	77.331	20	Li 56
				77.3167	19	Ge 57
			$C_2H_3N-K^{41}$	‡65.13	5	He 51
20	Ca	40	$C_3H_4-Ca^{40}$	68.7341	15	Ge 57
		42	$C_3H_6-Ca^{42}$	88.3500	22	Ge 57
		43	$C_3H_7-Ca^{43}$	96.0186	26	Ge 57
		44	CO_2-Ca^{44}	34.3442	24	Ge 57
		46	CSH_2-Ca^{46}	34.0462	39	Ge 57
		48	C_4-Ca^{48}	47.4964	55	Ge 57
21	Sc	45	$CSH-Sc^{45}$	23.9873	18	Ge 57
22	Ti	46	CSH_2-Ti^{46}	35.1026	14	Ge 57
		47	CSH_3-Ti^{47}	43.8035	30	Ge 57
		48	$SO-Ti^{48}$	19.0476	12	Ge 57
		49	C_4H-Ti^{49}	59.9781	15	Ge 57
		50	$C_4H_2-Ti^{50}$	70.8839	18	Ge 57
23	V	50	$C_4H_2-V^{50}$	68.5076	15	Ge 57
		51	$C_4H_3-V^{51}$	79.5223	18	Ge 57
24	Cr	50	$C_4H_2-Cr^{50}$	69.6218	18	Ge 57
		52	$C_4H_4-Cr^{52}$	90.8165	17	Ge 57
		53	$C_4H_5-Cr^{53}$	98.5062	21	Ge 57
24	Cr	54	$C_4H_6-Cr^{54}$	108.1099	23	Ge 57
25	Mn	55	$C_4H_7-Mn^{55}$	116.7547	22	Ge 57
			$\frac{1}{3}Ho^{165}-Mn^{55}$	§38.3	3	Ho 54
26	Fe	54	$C_4H_6-Fe^{54}$	107.374	4	Qu 56a
		56	$Si^{28}-\frac{1}{2}Fe^{56}$	§9.30	6	Du 50
			$C_4H_8-Fe^{56}$	127.698	4	Qu 56a
		57	$C_4H_9-Fe^{57}$	135.005	7	Qu 56a
		58	$C_4H_{10}-Fe^{58}$	144.977	4	Qu 56a
27	Co	59	$C_2H_3O_2-Co^{59}$	80.1466	23	Qu 56a
28	Ni	58	$C_4H_{10}-Ni^{58}$	142.941	7	Qu 56a
			$C_3H_6O-Ni^{58}$	106.52	15	Ea 56
		60	$C_2H_4O_2-Ni^{60}$	90.387	6	Qu 56a
				90.82	15	Ea 56
		61	$C_2H_5O_2-Ni^{61}$	97.894	5	Qu 56a
		62	$C_5H_2-Ni^{62}$	87.339	6	Qu 56a
		64	SO_2-Ni^{64}	33.901	5	Qu 56a
29	Cu	63	$C_5H_3-Cu^{63}$	93.909	7	Qu 56a
		65	$C_5H_5-Cu^{65}$	111.377	4	Qu 56a
30	Zn	64	$O_2-\frac{1}{2}Zn^{64}$	25.45	15	Kr 55
				25.2633	26	Qu 56a
			SO_2-Zn^{64}	32.7687	32	Qu 56a
		66	$C_5H_6-Zn^{66}$	120.935	6	Qu 56a
			$\frac{1}{2}Xe^{132}-Zn^{66}$	25.61	15	Kr 55
30	Zn	67	$C_5H_7-Zn^{67}$	127.675	7	Qu 56a
			$\frac{1}{2}Xe^{134}-Zn^{67}$	25.25	20	Kr 55
		68	$C_5H_8-Zn^{68}$	137.781	4	Qu 56a
			$\frac{1}{2}Xe^{136}-Zn^{68}$	27.70	20	Kr 55
		70	$C_5H_{10}-Zn^{70}$	152.953	6	Qu 56a
36	Kr	84	$CH_2Cl_2^{35}-Kr^{84}$	41.849	33	Kr 57
		86	$\frac{1}{3}Xe^{129}-\frac{1}{2}Kr^{86}$	12.969	17	Kr 57
			$CHCl^{35}F_2^{19}-Kr^{86}$	62.733	49	Kr 57
38	Sr	88	$CO_2-\frac{1}{2}Sr^{88}$	§37.00	18	Du 51a
39	Y	89	$C_8H_9-Y^{89}O$	§169.84	11	Co 54
54	Xe	126	$C_5H_3-\frac{1}{2}Xe^{126}$	§71.27	7	Ha 52
		129	$\frac{1}{3}Xe^{129}-\frac{1}{2}Kr^{86}$	12.969	17	Kr 57
		130	$C_6H_{10}O_3-Xe^{130}$	159.53	3	Jo 57
		131	$C_{10}H_{11}-Xe^{131}$	181.05	4	Jo 57
		132	$CO_2-\frac{1}{3}Xe^{132}$	21.762	20	Kr 57
			$\frac{1}{2}Xe^{132}-Zn^{66}$	25.61	15	Kr 55
			$C_{10}H_{12}-Xe^{132}$	189.79	5	Jo 57
		134	$\frac{1}{2}Xe^{134}-Zn^{67}$	25.25	20	Kr 55
			$C_{10}H_{14}-Xe^{134}$	204.20	5	Jo 57
		136	$\frac{1}{2}Xe^{136}-Zn^{68}$	27.70	20	Kr 55
			$C_{10}H_{16}-Xe^{136}$	218.055	25	Jo 57
55	Cs	133	$C_{10}H_{13}-Cs^{133}$	196.66	7	Jo 57
56	Ba	130	$C_6H_{10}O_3-Ba^{130}$	156.24	20	Jo 57
		132	$C_{10}H_{12}-Ba^{132}$	188.84	12	Jo 57
		134	$C_{10}H_{14}-Ba^{134}$	205.36	8	Jo 57
		135	$C^{13}C_9H_{14}-Ba^{135}$	207.40	10	Jo 57
		136	$C_{10}H_{16}-Ba^{136}$	220.89	9	Jo 57
		137	$C^{13}C_9H_{16}-Ba^{137}$	223.08	6	Jo 57
		138	$C_{10}H_{18}-Ba^{138}$	236.03	8	Jo 57
57	La	138	$C_{10}H_{18}-La^{138}$	234.17	20	Jo 57
		139	$C^{13}C_9H_{18}-La^{139}$	238.23	6	Jo 57
58	Ce	136	$C_{10}H_{16}-Ce^{136}$	218.19	20	Jo 57
		138	$C_{10}H_{18}-Ce^{138}$	234.89	20	Jo 57
		140	$C_{10}H_{20}-Ce^{140}$	251.29	6	Jo 57
		142	$C_{10}H_{22}-Ce^{142}$	262.93	7	Jo 57

regularly increase in width as heavier atoms are investigated. If this latter had not been the case, a systematic error that increased with doublet spacing might have provided an explanation for the mounting inconsistencies.

Among heavier atoms, particularly in the iron-nickel-zinc region, certain large discrepancies[15] between mass differences as calculated from mass spectroscopy and as derived from reaction energies have been virtually removed.[16] Large inconsistencies still exist in the 50-neutron region.

In short, the absolute agreement between the results from these two major sources of atomic mass information has much improved in the past three years. The remaining dissonance would be greatly lessened and, incidentally, great joy would be brought to the heart of the mass spectroscopist, if a *small* systematic error were to be found among Q-value determinations.

IV. TABLE OF ATOMIC MASS DIFFERENCES

In Table II are listed the atomic mass differences which have been obtained by mass spectroscopic methods in the past three years. These data supplement those in Tables II, III, and IV of reference 1. Appearing herein is a good deal of information which has not, at the time of writing (June 1957), been published.

In reference 1 the doublets that were used to obtain the masses of the secondary standards H^1, D^2, and C^{12} were tabulated separately. Recently, however, several new doublet cycles that provide means of calculating these masses have been introduced,[17,6,18] with the result that there is now a rather formidable number of "fundamental" doublets, some of which are sacrosanct in one laboratory but not in another. For this reason, and also because a number of heavier atoms are likely to be added[19] to the list of secondary standards, all the mass data are here shown in a single table.

The mass differences in Table II are arranged according to element in order of increasing atomic number, Z, and, within the given element, according to isotope in order of increasing mass number, A. The entries associated with any particular nuclide represent those data which may be useful in calculating the mass of the nuclide in question. As a result there are many double entries, for example, the H_2O-DO mass difference is listed under both H and D. On the other hand, the $C_{12}H_{10}-Sm^{154}$ mass difference is listed only under Sm^{154}. Although it does involve both H and C, this difference is not of practical use in determining the masses of these two atoms.

[15] Kerr, Taylor, and Duckworth, Nature **176**, 458 (1955).
[16] Quisenberry, Scolman, and Nier, Phys. Rev. **104**, 461 (1956).
[17] M. E. Kettner, Phys. Rev. **102**, 1065 (1956).
[18] L. G. Smith, Bull. Am. Phys. Soc. Ser. II, **2**, 223 (1957).
[19] H. E. Duckworth, Progr. in Nuclear Phys. **6**, 139 (1957).

TABLE II.—*Continued.*

Element Z	A	Doublet	ΔM in mMU	Error	Reference
59 Pr	141	$C_{11}H_9-Pr^{141}$	163.00	3	Jo 57
60 Nd	142	$C_{10}H_{22}-Nd^{142}$	264.74	3	Jo 57
	143	$C^{13}C_{10}H_{10}-Nd^{143}$	172.08	10	Jo 57
	144	$C_{10}H_5F^{19}-Nd^{144}$	127.77	7	Jo 57
	145	$C_{10}H_6F^{19}-Nd^{145}$	133.33	19	Jo 57
	146	$C_{10}H_7F^{19}-Nd^{146}$	140.53	6	Jo 57
	148	$C_2{}^{13}C_8H_7F^{19}-Nd^{148}$	143.46	6	Jo 57
	150	$C_9H_{10}O_2-Nd^{150}$	147.30	7	Jo 57
62 Sm	144	$C_{10}H_5F^{19}-Sm^{144}$	125.92	9	Jo 57
	147	$C^{13}C_9H_7F^{19}-Sm^{147}$	142.09	3	Jo 57
	148	$Sm^{148}-Sm^{147}$	†1000.25		Jo 57
		$Sm^{149}-Sm^{148}$	†1002.71		Jo 57
	149	$Sm^{149}-Sm^{148}$	†1002.71		Jo 57
		$Sm^{150}-Sm^{149}$	†1000.42		Jo 57
	150	$C_9H_{10}O_2-Sm^{150}$	151.23	7	Jo 57
	152	$C_{12}H_8-Sm^{152}$	143.29	13	Jo 57
	154	$C_{12}H_{10}-Sm^{154}$	156.37	15	Jo 57
63 Eu	151	$C_{12}H_7-Eu^{151}$	135.26	17	Jo 57
		$C_{13}H_{11}-Eu^{151}O$	171.69	19	Jo 57
	153	$C^{13}C_{12}H_{12}-Eu^{153}O$	181.8	4	Jo 57
64 Gd	154	$Gd^{155}-Gd^{154}$	†1002.15	6	Jo 57a
	155	$Gd^{156}-Gd^{155}$	†999.90	6	Jo 57a
	156	$Gd^{157}-Gd^{156}$	†1002.20	6	Jo 57a
	157	$Gd^{158}-Gd^{157}$	†1000.53	6	Jo 57a
66 Dy	160	$Dy^{161}-Dy^{160}$	†1002.10	6	Jo 57a
	161	$Dy^{162}-Dy^{161}$	†1000.21	6	Jo 57a
	162	$Dy^{163}-Dy^{162}$	†1002.26	6	Jo 57a
	163	$Dy^{164}-Dy^{163}$	†1000.80	6	Jo 57a
67 Ho	165	$\frac{1}{3}Ho^{165}-Mn^{55}$	§38.3	3	Ho 54
68 Er	166	$Er^{167}-Er^{166}$	†1002.06	6	Jo 57a
	167	$Er^{168}-Er^{167}$	†1000.65	6	Jo 57a
70 Yb	170	$Yb^{171}-Yb^{170}$	†1001.88	6	Jo 57a
	171	$Yb^{172}-Yb^{171}$	†1000.40	6	Jo 57a
	172	$Yb^{173}-Yb^{172}$	†1002.17	6	Jo 57a
	173	$Yb^{174}-Yb^{173}$	†1000.97	6	Jo 57a
72 Hf	176	$Hf^{177}-Hf^{176}$	†1002.25	6	Jo 57a
	177	$Hf^{178}-Hf^{177}$	†1000.88	6	Jo 57a
	178	$Hf^{179}-Hf^{178}$	†1002.36	6	Jo 57a
	179	$Hf^{180}-Hf^{179}$	†1001.13	6	Jo 57a
74 W	182	$W^{183}-W^{182}$	†1002.23	6	Jo 57a
	183	$W^{184}-W^{183}$	†1000.99	6	Jo 57a
76 Os	186	$Os^{187}-Os^{186}$	†1002.14	6	Jo 57a
	187	$Os^{188}-Os^{187}$	†1000.33	6	Jo 57a
	188	$Os^{189}-Os^{188}$	†1002.55	6	Jo 57a
	189	$Os^{190}-Os^{189}$	†1000.52	6	Jo 57a
78 Pt	194	$Pt^{195}-Pt^{194}$	†1002.45	6	Jo 57a
	195	$Pt^{196}-Pt^{195}$	†1000.49	6	Jo 57a
80 Hg	198	$Hg^{199}-Hg^{198}$	†1001.82	6	Jo 57a
	199	$Hg^{200}-Hg^{199}$	†1000.31	6	Jo 57a
	200	$Hg^{201}-Hg^{200}$	†1002.26	6	Jo 57a
	201	$Hg^{202}-Hg^{201}$	†1000.64	6	Jo 57a
82 Pb	206	$Pb^{207}-Pb^{206}$	†1001.74	6	Jo 57a
	207	$Pb^{208}-Pb^{207}$	†1001.06	6	Jo 57a

Symbols appearing in Table II have the following significance: ‡ indicates a datum omitted in the tables of reference 1, § indicates a correction to the tables of

reference 1, and † designates mass differences which are not true doublets but are, rather, mass-unit differences of the type described in the concluding paragraph of Sec. II. A datum in parentheses indicates that the investigator originally responsible for it now views it with suspicion. As before, it should be assumed that later values supersede earlier values from the same laboratory.

The reader should consult the original papers for descriptions of the assigned errors. In most cases these are probable errors based *only* on the internal consistency of the data. The frequency with which two entries for the same doublet agree within the stated errors suggests that systematic errors are roughly comparable in size to the statistical ones.

V. ACKNOWLEDGMENTS

I have been greatly helped in the preparation of this review by a number of persons who have kindly provided me with pre-publication information concerning their work and who have checked the accuracy of the first draft of Table II. These include J. L. Benson, V. B. Bhanot, H. Ewald, C. F. Giese, W. H. Johnson, J. T. Kerr, M. E. Kettner, J. Mattauch, A. O. Nier, K. Ogata, K. S. Quisenberry, T. T. Scolman, and L. G. Smith. I also much appreciate the assistance of Gertrude Isenor who has compiled most of the information in Table II and has assiduously checked it.

VI. REFERENCES

Co 54 Collins, Johnson, and Nier, Phys. Rev. **94**, 398 (1954).
De 56 Demirkhanov, Gutkin, Dorokhov, and Rudenko, Atomnaya Energ. **2**, 21 (1956); J. Nuclear Energy **3**, 251 (1956).
Du 50 Duckworth, Johnson, Preston, and Woodcock, Phys. Rev. **78**, 386 (1950).
Du 51a H. E. Duckworth and R. S. Preston, Phys. Rev. **82**, 468 (1951).
Ea 56 Eastman, Isenor, Bainbridge, and Duckworth, Phys. Rev. **103**, 145 (1956).
Ge 57 C. F. Giese and J. L. Benson, Bull. Am. Phys. Soc. Ser. II, **2**, 223 (1957).
Ha 52 R. E. Halsted, Phys. Rev. **88**, 666 (1952).
Ho 54 B. G. Hogg and H. E. Duckworth, Can. J. Phys. **32**, 65 (1954).
He 51 A. Henglein, Z. Naturforsch. **6a**, 745 (1951).
Jo 57 W. H. Johnson, Jr., and A. O. Nier, Phys. Rev. **105**, 1014 (1957).
Jo 57a W. H. Johnson, Jr., and V. B. Bhanot, Bull. Am. Phys. Soc. Ser. II, **2**, 224 (1957).
Ket 56 M. E. Kettner, Phys. Rev. **102**, 1065 (1956).
Kr 55 Kerr, Isenor, and Duckworth, Z. Naturforsch **10a**, 840 (1955).
Li 56 H. Liebl and H. Ewald, Z. Naturforsch **11a**, 406 (1956).
Ma 54 J. Mattauch and R. Bieri, Z. Naturforsch. **9a**, 303 (1954).
Qu 56 Quisenberry, Scolman, and Nier, Phys. Rev. **102**, 1071 (1956).
Qu 56a Quisenberry, Scolman, and Nier, Phys. Rev. **104**, 461 (1956).
Qu 57 Quisenberry, Giese, and Benson, Bull. Am. Phys. Soc. Ser. II, **2**, 223 (1957).
Sch 56 Schierstedt, Ewald, Liebl, and Sauermann, Z. Naturforsch **11a**, 216 (1956).
Sco 56 Scolman, Quisenberry, and Nier, Phys. Rev. **102**, 1076 (1956).
Sm 57 L. G. Smith, Bull. Am. Phys. Soc. Ser. II, **2**, 223 (1957).

Table of Total Beta-Disintegration Energies. II.*†

L. J. LIDOFSKY

Columbia University, New York, New York

INTRODUCTION

THE "Table of Total Beta-Disintegration Energies,"[1] which appeared in 1954, has been used widely as a data summary and as a source for evaluating mass differences.[2] Since then, a great deal of new data has appeared in the literature, enough, in fact, to warrant the issuance of this report as a supplement to (I). Information available to the author up to about March 1, 1957, is included.

EXPLANATION OF TABLE

For the purpose of continuity, the symbols and format used in (I) have been retained. As before, in the case of a complex decay, the Q value is ordinarily evaluated from the single branch which is best known. However, in those cases where several branches have been investigated by coincidence-spectrometer methods, each is considered in making the final evaluation. Also, in certain cases where the decay scheme is not well established from the information on beta decay, information on the position of levels from nuclear reactions is used as a guide.

The adopted Q value is obtained by weighting the various measurements inversely as the squares of their errors. In those cases where the author did not quote a quantitative error, an estimate of the error was made. The estimates employed were in general based on criteria similar to those indicated in (I). However, in cases where there were indications of the accuracy of the experiment, errors were estimated in accord with those indications.

ACKNOWLEDGMENTS

The author wishes to acknowledge the invaluable assistance afforded by several data summaries and reviews which have appeared in the literature since 1954. These publications offered a check on the completeness of abstracting and, in addition, provided access to certain data not otherwise available to the author. Among these publications are

(i) The nuclear data summaries of Nuclear Science Abstracts which appear annually as No. 24-B of the volume for that year.

(ii) Way, King, McGinnis, and van Lieshout, "Nuclear level schemes," TID-5300 (U. S. Atomic Energy Commission, September, 1955).

(iii) F. Ajzenberg-Selove, " 'Classical' nuclear physics in the USSR," Nuovo cimento Suppl. 4, 2 (1956).

* Prepared at the suggestion of the Subcommittee on Nuclear Constants of the Committee on Nuclear Science of the National Research Council as part of a program of the compilation of experimental data relating to atomic masses.

† This work was partially supported by the U. S. Atomic Energy Commission.

[1] R. W. King, Revs. Modern Phys. 26, 327 (1954), hereinafter referred to as (I).

[2] A. H. Wapstra, Physica 21, 367 (1955); Physica 21, 385 (1955). J. R. Huizenga, Physica 21, 410 (1955). Mattauch, Waldman, Bieri, and Everling, Z. Naturforsch. 11A, 525 (1956). Way, King, McGinnis, and van Lieshout, "Nuclear level schemes," TID-5300 (U. S. Atomic Energy Commission, September, 1955).

TABLE I.

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	%	METHOD	REF.	COMMENTS
$_{2\ 4}He \xrightarrow{6} {}_{3\ 3}Li$ 0^+ 2.9 $\underline{1^+}$	3.50	2	0.80^s	β_0^-	3.50	2		s	56S1	F-K plot linear to 0.250
$_{6\ 5}C \xrightarrow{11} {}_{5\ 6}B$ 3/2 3.2 $\underline{3/2^-}$	1.994	6	20.4^m	β_0^+	0.968	8		s	54W1	
$_{5\ 7}B \xrightarrow{12} {}_{6\ 6}C$ 1^+ 4.2 0^+	13.43	6	$.018^s$	β_0^-	13.37	20*		scin	56T1	
$_{5\ 8}B \xrightarrow{13} {}_{6\ 7}C$ $3/2^-$ $\underline{1/2^-}$	~ 13		$.035^s$	β^-				scin	56N1	β^- with E_{max} ~ E_{max} from B^{12} seen following $Li^7(Li^?,p)B^{13}$
$_{6\ 8}C \xrightarrow{14} {}_{7\ 7}N$ $\underline{0^+}$ 9.0 $\underline{1^+}$	0.156	1	$~5600^y$	β_0^-	0.1585	5		s	55P1	F-K plot linear
					0.156	2		pc	54M1	
					0.155	1		s	54F1	
					0.155	1		s	53W1	
					0.157	2		pc	53M1	
$_{8\ 6}O \xrightarrow{14} {}_{7\ 7}N$ 0^+ 3.5 0^+ $\underline{1^+}$	5.15	4	72^s	β^+	1.835	8*		s	54G1	β^+ decay to 2.3 level in N^{14}
					1.830	30		scin	54P1	
				β_0^+	(4.1)		~.6%		56S2	
$_{6\ 9}C \xrightarrow{15} {}_{7\ 8}N$ $5/2^+$ 3.9 $3/2^-$ 8.2 $\underline{1/2^-}$	9.7	3	2.3^s	β^-	4.3	1*	80%	scin	56D1	
				β_0^-	9.8	2*	20%	scin	56D1	
				γ	5.3	1*		scin	56D1	
$_{8\ 7}O \xrightarrow{15} {}_{7\ 8}N$ $1/2^-$ 3.9 $\underline{1/2^-}$	2.855	10	124^s	β_0^+	1.837	10		s	57K1	
					1.823	5		s	57K1	
$_{9\ 8}F \xrightarrow{17} {}_{8\ 9}O$ $5/2^+$ 3.4 $\underline{5/2^+}$	2.770	6	66^s	β_0^+	1.748	6		s	54W1	
$_{8\ 11}O \xrightarrow{19} {}_{9\ 10}F$ $5/2^+$ 5.6 $5/2^+$ $\underline{1/2^+}$	4.7	3	29.4^s	γ	0.200	2		scin	54J1	4.5 β^- to 0.200 level in F^{19} (54J1)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Ne 10 13 --23--> Na 11 12	4.39	4	38^s	β_0^-	4.39	5	67	scin	57P1	(β 3.95)(γ 0.436) confirm decay scheme (57P1)
$5/2^+$ $5/2^+$				β^-	3.9	3		scin	56G1	
5.3 $3/2^+$					3.95	5	32	scin	57P1	
				γ	.436	4		scin	57P1	
					.440	5		scin	56G1	
Ne 10 14 --24--> Na 11 13	2.45	4	3.38^m	β^-	1.98	5	92	scin	56D2	(β 1.98)(γ 0.472) and (β 1.10)(γ 0.472)(γ 0.878) observed (56D2)
0^+ 4.4 1^+				β^-	1.10	5	8	scin	56D2	
1^+				γ	.472	5		scin	56D2	
4^+				γ	.878	5		scin	56D2	
Na 11 13 --24--> Mg 12 12	5.512	5	14.9^h	β^-	1.400	15		s	55D1	
4^+ 6.1 4^+										
12.7 2^+										
0^+										
Al 13 11 --24--> Mg 12 12	14	1	2.1^s	β^+ ~ 8.5			weak	scin	55G1	β^+ presumably goes to ~4.22 mev state (55G1)
4^+ ? >4.1				several γ rays observed				55G1		
0^+										
Na 11 14 --25--> Mg 12 13	3.8	2	62^s	β_0^-	4.0	2	65	scin	55M1	(β)(γ) confirm ~ 3.8 β^- as ground state transition
$5/2^+$					3.8	2		scin	54M3	
$1/2^+$					3.65	25		a	54N1	
5.3 $5/2^+$				γ	0.576	10		scin	55M1	
				other γ's				scin	55H1	
Al 13 13 --26--> Mg 12 14	2.99	10	$\sim10^{6y}$	β^+	1.13	6		scin	57J1	(β^+)(1.8γ) and ($\gamma\pm$)(1.8γ) observed (55L1), (55H1).
5^+					1.17	5		scin	55L1	ΔJ = 3, no shape (55L1),(57J1)
>14 2^+					1.30	15		a	55H1	
0^+					~1			a	54S1	
				γ	1.84	10		scin	57J1	
					1.76	10		scin	55L1	
					1.82	10		scin	55H1	
					1.9	1		scin	54S1	
Mg 12 15 --27--> Al 13 14	2.595	10	9.5^m	β_1^-	1.78	1		s	54M2	(β_1)(γ_1) and (β_2)(γ_2) (56L1), (54M2)
$1/2^+$ 4.8 $3/2^+$				β_2	1.592	22		s	54D1	
4.8 $1/2^+$					1.5	1		scin	54M2	
$5/2^+$				γ_1	0.834	10		scin	56L1	
					0.834	8		s	54D1	
				γ_2	1.015	10		scin	56L1	
					1.015	7		s	54D1	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Si --27--> Al 14 13 13 14 5/2+ 3.6 5/2+	4.83	4	4.3^s	β_0^+	3.82	4		scin	55R1	
Mg --28--> Al 12 16 13 15 0+ 4.9 1+ 2+ 2+ 3+	1.837	5	21.4^h	γ_1	.0319	1		s	54S2	$(\gamma)(\gamma)$ and γ- well crystal summing indicate all γ's previously reported are coincident with 0.032 γ. (54S2)
				γ_2	1.346	1		s	54S2	
				and other γ's						
Al --28--> Si 13 15 14 14 3+ 4.9 2+ 0+	4.65	1		β^-	2.75	10		a	54N1	$(\beta)(\gamma)$ observed (54N1), (54S2)
					~3			a	54S2	
				γ	1.83	5		scin	54N1	
					1.76	2		scin	54S2	
P --28--> Si 15 13 14 14 2, 3+ >4.7 2+ 0+	13.4	4	0.28^s	β^+	10.6	4	50	scin	55G1	other γ's and lower energy β^+ present (55G1)
				γ	1.79	2	strong	scin	55G1	
Al --29--> Si 13 16 14 15 5/2+ 4.9 3/2+ 3/2+ 1/2+	3.95	10	6.6^m	β_1^-	1.55	10*	~25	a	54N2	$(\beta\ 1.55)(\gamma\ 2.42)$ and $(\beta\ 2.65)(\gamma\ 1.31)$ (54N2)
				β_2^-	2.65	10*		a	54N2	
				γ_1	2.43	5*		scin	56B1	
					2.43	5*	9	scin	55R2	
					2.42	5		scin	54N2	
				γ_2	1.28	2*		scin	56B1	
					1.28	2*		scin	55R2	
					1.31	2		scin	54N2	
P --29--> Si 15 14 14 15 1/2+ 3.7 1/2+	4.96	1	4.5^s	β_0^+	3.82	4		scin	55R1	$(\beta^+)(\gamma)$ studies confirm this as ground state transition (55R1)
					3.945	5		s	55R2	
P --30--> Si 15 15 14 16 1+ 5.0 0+	4.28	5	2.52^m	β_0^+	3.24	4		s	56G1	
				No γ					56G1	
S --31--> P 16 15 15 16 1/2+ 3.7 1/2+	5.44	3	2.5^s	β_0^+	4.42	3		scin	55R1	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
P -32-> S 15 17 16 16 1^+ 7.9 0^+	1.706	3	14.3^d	β_0^-	1.704	10		s	56A1	
					1.712	5		s	56P1	
					1.712	8		s	56A2	
					1.700	5		s	56A2	
					1.711	6		s	54D1	
Cl -32-> S 17 15 16 16 1^+ >4.6 0^+	≥12.8	4	0.35′	β^+	9.5	4	50	scin	56G1	assume β^+ decay to 2.25 level in S^{32}
P -33-> S 15 88 16 17 $1/2^+$ 5.0 $3/2^+$	0.249	2	25d	β^-	0.250	5		s	54E1	
					0.249	2		s	54N3	
A -34-> Cl 18 16 17 17 0^+ 0^+	> 0		$0.2^{-}2^s$	β^+					55K1	identified by Cl^{34} daughter
S -35-> Cl 16 19 17 18 $3/2^+$ 5.0 $3/2^+$	0.1670	5	87.1^d	β_0^-	0.165	5		s	54F2	F-K plot linear to 5 kev (54M1)
					0.167	2		pc	54M1	
A -35-> Cl 18 17 17 18 $3/2^+$ 3.8 $3/2^+$	5.98	4	1.83^s	β^+	4.96	4	93	s	56K1	weak γ rays (1.19 + 1.73) also reported (56K1)
Cl -36-> S 17 19 16 20 2^+ 0^+	> 0		$3.1\times10^5 y$	ϵ_K			1.7	scin	55D2	From relative intensity of (56J1) and K auger electrons/β-
								pc		
								s		
Cl -36-> A 17 19 18 18 2^+ 13.5 0^+	0.714	5	$3.1\times10^5 y$	β^-	0.714	4		scin	56J1	Shape is fitted by ΔJ = 2,no. (56J)
S --37-> Cl 16 21 17 20 $7/2^-$ 4.3 $5/2^-$ or $7/2^-$ 9.2[1] $3/2^+$	4.7	1	5.0^m	β^-	1.6	1*	90	scin	56M1	10 per cent ~4.7 β also reported (56M1)
				γ	3.1	1*		scin	56M1	
					3.09	3		scin	56S3	
A --37-> Cl 18 19 17 20 $3/2^+$ 5.1 $3/2^+$	0.815	2	34.1^d	continuous γ spectrum						Electron capture
				E_ϵ	0.812	8		s recoil	55K2	excess low energy photons supports Glauber-Martin Theory (56S4), (55L2).
					0.814	2		s recoil	55S1	
				E_γ max	0.820	10		scin	56S4	K Binding energy = 0.003
					0.820	10		scin	55L2	
					0.815	20		scin	54S3	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
K —37→ A 19 18 18 19 3/2$^+$ 3.5 3/2$^+$	6.1	1	1.2^s	β_o^+	5.1	1*		s	56S5	
Cl —39→ A 17 22 18 21 3/2$^+$ 5.6 3/2$^+$ 3/2, 5/2$^-$ 8.3[1] 7/2$^-$	3.43	2	56^m	β_o^-	3.43	2	7	s	56P2	$(\beta_1)(\gamma_1)$, and $(\beta_1)(\gamma\ 0.246)(\gamma\ 1.270)$
				β_1^-	1.90	5	85	scin$\beta\gamma$	56P2	give good energy match (56P2)
										ΔJ = 2,yes observed for β_o(56P2)
				γ_1	1.520				56P2	
Ca —39→ K 20 19 19 20 3/2$^+$ 3.8 3/2$^+$	6.6	1	0.95^s	β_o^+	5.58	4		scin	55R1	
Cl —40→ A 17 23 18 22 2$^-$ ≥9.6[1] 0$^+$	7.5	5	1.4^m	β^-	~7.5		weak	scin	56M2	Strong ~3.0 β component is assumed to decay to a (-) parity state at ~4.2 mev (56M2)
K —40→ Ca 19 21 20 20 4$^-$ 18.1 0$^+$	1.33	1	1.3x10^{9y}	β_o^-	1.320	20		scin	55K3	ΔJ = 4,yes shape.
A —41→ K 18 23 19 22 7/2$^-$ 4.8 7/2$^-$ 8.4[1] 3/2$^+$	2.49	1	1.78^h	β^-	1.199	8	99.1	s	56S6	~2.48 mev ground state decay spectrum can be fit to ΔJ = 2,yes shape (56S6)
				γ	1.290	5		scin	56S6	
					1.298	10		scin	55K4	
K —42→ Ca 19 23 20 22 2$^-$ 7.5 2$^+$ 8.4[1] 0$^+$	3.55	1	12.5^h	β_o^-	3.545	10	82	s	56P3	β_o has ΔJ = 2,yes shape (56P3, 54K1)
					3.560	30	82	s	54K1	
				β_1^-	1.985	10	18	s$\beta\gamma$	56P3	
					1.970	20	18	s$\beta\gamma$	54K1	
				γ	1.530	10		s	56P3	
Sc —42?→ Ca 21 21 20 22 0$^+$ 0$^+$	>4		0.62^s	β^+	several mev			scin	55M1	K + α bombardment. Identified by yield. (55M1)
K —43→ Ca 19 24 20 23 3/2$^+$ 5.5 3/2$^+$ 3/2$^-$ 5/2$^-$ 8.8[1] 7/2$^-$	1.815	15	22^h	β_o^-	1.815	30	8	s	57B1	$(0.371\gamma)(0.614\gamma)$ and $(0.591\gamma)(0.314\gamma)$ indicate state at 0.985 fed by 0.82β ($\beta\gamma$), (57B1)
				β_1^-	0.825	10	87	s	57B1	β_o has ΔJ = 2,yes shape (57B1)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Sc -43-> Ca 21 22 20 23 $7/2^-$ 5.0 $7/2^-$	2.22	1	3.9^h	β^+_0	1.20	1	79	s	54L1	$(0.82\beta^+)(0.38\gamma)$ gives energy match (54L1)
Sc -44-> Ca 21 23 20 24 3^+, 3^+ 5.3 2^+ 0^+	3.65	1	4.0^h	β^+	1.471	5		sβγ	55B1	
				γ	1.159	3		s	55B1	
Sc -46-> Ti 21 25 22 24 4^+ 6.2 4^+ 13.1 2^+ 0^+	2.364	5	85^d	β^-	1.475	6	.0036	s	56W1	β^- has ΔJ = 2, no shape. (56W1)
				γ	0.892	3		scin	56J2	
					1.118	3			56J2	
Ca -47-> Sc 20 27 21 26 $7/2^-$ 8.5 $7/2^-$	2.00	4	4.7^d	β^-_0	1.94	2	83	s	56L2	F-K plot linear. Energy match provided by 0.66 β^- + 1.31γ (56L2, 55L3)
					1.9	2		a	55L3	
Sc -47-> Ti 21 26 22 25 $7/2^-$ 5.3 $7/2^-$ 6.1 $5/2^-$	0.605	5	3.44^d	β^-_0	0.610	5	26	s	56L2	Energy match by (0.44β)(0.160γ) (56L2, 56G2, 55N1, 55L4)
					0.600	10	40	s	56G2	
					0.596	10	36	s	55N1	
					0.62	3		a	55L4	
V -48-> Ti 23 25 22 26 4^+ ~ 6.2 4^+ 2^+ 0^+	4.02	2	16.1^d	β^+	0.694	6		s	55K5	β^+ decays to 2.32 state in Ti^{48}
Cr -48-> V 24 24 23 25 0^+ ≥ 4.1 1^+ 2^+ 4^+	1.4	2	24^h	γ_1	.116			scin	55V1	(0.12γ)(0.31γ) confirmed (55V1, 55W1) Q value selected on basis of < 2 per cent β^+ and on ft for an allowed transition.
					.118			scin	55W1	
				γ_2	.305			scin	55V1	
					.307			scin	55W1	
				β^+	< 2 per cent		$\gamma\pm$	scin	55V1	
							$\gamma\pm$	scin	55W1	
Ca -49-> Sc 20 29 21 28 $3/2^-$ 5.8 $3/2^-$ $7/2^-$	5.08	6	8.8^m	β^-	2.12	10		aβγ	56M3	(2.1β)(3.1γ) confirmed (56M3)
					1.95	5	8.8	scin	56O1	
					2.0	2		scin	55M2	
				γ	3.07	5		scin	56M3	
					3.10	3		scin	56O1	
					3.00	15		scin	55M2	
Sc -49-> Ti 21 28 22 27 $7/2^-$ 5.8 $7/2^-$	2.0	1	57.2^m	β^-	1.8	1		a	56M3	no γ (56O1)
					2.05	5	100	scin	56O1	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
V —49→ Ti										
23 26 22 27	0.62	1	330^d	continuous γ spectrum						K binding energy = 0.0055
7/2⁻ 6.1 7/2⁻				E_{max}	0.616	10		scin	56H1	γ spectrum fit by Glauber-Martin
					0.612	20		scin	54P2	theory
Sc —50→ Ti										
21 29 22 28	6.3	5	1.7^m	β⁻	~ 3.5		1†	scin	55M3	(1.17γ)(1.59γ) observed in crystal
5⁻, 4⁻(?) ≥ 5.1 4⁺				γ1	1.17	5	1†	scin	55M3	summing (55M3).
2⁺				γ2	1.59	10	1†	scin	55M3	approximately equal intensities
0⁺										indicate that all radiations may be in cascade.
Ti —51→ V										
22 29 23 28	2.46	4	5.8^m	β⁻	2.3	2	95	aβγ	56J3	F-K plot linear (55B2, 55M4)
3/2⁻ 3/2⁻					2.17	4		scin	55S2	(2.2β)(0.3γ) observed (56J3, 55S2,
4.9 5/2⁻					2.13	3		scin	55B2	55B2, 55M4)
7/2⁻					2.13	3		s	55M4	
					2.22	7		scin	53N1	
				γ	0.325	5*		scin	55S2	(1.5β)(0.93γ) gives energy match
					0.323	2*		scin	55B2	(55B2, 56J3)
					0.32	1		scin	55M4	
Cr —51→ V										
24 27 23 28	0.750	5	27^d	ε						K binding energy = 0.006
7/2⁻ 5.4 7/2⁻				continuous γ spectrum						
				E_γ (max)	0.746	22		scin	56V1	
					0.750	5	93	scin	55B3	
					0.774	50		scin	55C1	
V —52→ Cr										
23 29 24 28	4.13	10	3.77^m	β⁻	2.73	6		scin	55B4	
2⁺ 5.5 2⁺				γ	1.49	4		scin	55B4	
0⁺										
Mn —52→ Cr										
25 27 24 28	4.735	25	5.7^d	β^+_m see comment						
2⁺ see comment 2⁺					2.631	15		s	56A1	decay is from metastable 0.39 state in Mn^{52} to 1.47 state in Cr^{52}.
6⁺ 0⁺										
Fe —52→ Mn										
26 26 25 27	≥2.22	2	7.8^h	β⁺	0.804	10		s	56A1	unobserved low energy γ in Mn^{52} is probable.
0⁺ 4.7 21^m 2⁺										
0⁺										
V —53→ Cr										
23 30 24 29	≥2.53	5	1.7^m	β⁻	2.53	5		scin	56S7	23^h activity previously assigned here not V^{53} (56S7, 55S3).
? ?				γ	1.00	1	st	scin	56S7	
3/2⁻					1.29		wk	scin	56S7	β systematics suggest Q ~ 2.5 (54W2)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Mn $\xrightarrow{53}$ Cr 25 28 24 29 7/2$^-$ (?) 3/2$^-$	>0		14C^{y}	k xrays observed					55W2	
V $\xrightarrow{54}$ Cr 23 31 24 30 (5$^+$) ≥5.3 2$^+$ 0$^+$	≥5.1	2	55^{s}	β^- γ_1 γ_2 γ_3	3.3 0.835 0.990 2.21	2 9 10 4	 weak	scin scin scin scin	56S7 56S7 56S7 56S7	(0.835γ)(0.990γ)(E_γ>1) observed. 0.835γ more intense than 0.99γ. β decay systematics suggest Q ~ 6-7 (54W2)
Mn $\xrightarrow{54}$ Cr 25 29 24 30 3$^+$ 6.1 2$^+$ 0$^+$	1.36	2	291^{d}	ϵ continuous γ-spectrum E_γ max γ	 0.522 0.840	 20 7		 (x-cont.γ)scin scin	 56J4 54M3	K binding energy .006
Mn $\xrightarrow{54}$ Fe 25 29 2 28 3$^+$ 0$^+$	see comment		291^{d}				~ 10^{-3} (see comment)			~ 1 mev electrons assigned here previously are probably e$^-$ from 0.84γ (see Mn^{54}---->Cr^{54}). Probable assignment of 3$^+$ to Mn^{54} from aligned nucleus studies (54G2) would suggest β^-/ϵ ~ 10^{-3} for a 1 mev β^- transition. (55T1).
Cr $\xrightarrow{55}$ Mn 24 31 25 30 3/2$^-$ 5.1 5/2$^-$	2.82	14	3.6^{m}	β^- ~ noγ	2.80	14*	 <.01	a	54B1 56H2	
Fe $\xrightarrow{55}$ Mn 26 29 25 30 3/2$^-$ 5/2$^-$	0.220	4	2.9^{y}	ϵ continuous γ spectrum $E_{\gamma max}$	 0.213 0.223 0.225	 10 10 10		 scin scin scin	 56S4 54M4 54S3	K binding energy = 0.0071
Co $\xrightarrow{55}$ Fe 27 28 26 29 7/2$^-$ 6.4 5/2$^-$,7/2$^-$ 3/2$^-$	3.46	1	18.2^{h}	β^+ γ	1.500 0.935	8 9*		s s	54C1 54C1	(1.03β^+)(1.41γ) gives good energy match (54C1).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Co $\xrightarrow{56}$ Fe 27 29 26 30	4.610	15	77^d	β^+	1.47 1.496	3 15$^\times$		s s	56K2 55H2	Many other γ's present disagreement on presence and nature of lower energy β^+ groups. Decay scheme including spin assignments established by $(\gamma)(\gamma)$ studies.
4^+ 8.5 4^+ 2^+				γ	0.845 0.85 0.845	8* 1* 5*		scin scin s	56K2 56C1 55H2	
0^+				γ_2	1.24 1.20 1.238	1* 1* 5*		scin scin s	56K2 56C1 55H2	
Ni $\xrightarrow{56}$ Co 28 28 27 29 0^+ 4^+	> 1.75		6.4^d	γ	1.74	5	wk	scin	55W7	Many lower energy γ's observed. β systematics suggest Q ~ 2.7 (54W2)
Mn $\xrightarrow{57}$ Fe 25 32 26 31	2.7	2	1.7^m	β^-	2.6	2		a scin	54C2	Weak γ's present (54C2). Same γ's appear in Co $\xrightarrow{57}$ Fe^{57}
5/2, $7/2^-$ ≥ 4.7 $5/2^-$				γ_1	0.134	6	str	scin	54C2	
$1/2^-$, $3/2^-$				γ_2 see comments	0.117	4	str	scin	54C2	
Co $\xrightarrow{57}$ Fe 27 30 26 31 $7/2^-$ 6.0 $5/2^-$	0.57	3	270^d	continuous γ spectrum E_γ max	0.426	30	(x)(cont.γ)	scin	56J4	β^+ previously assigned here not Co^{57}. (55M5, 55C2, 55L5, 55M6, 55C3) K binding energy 0.008
$3/2^-$, $1/2^-$				γ_1	0.137 0.1374 0.13631	2 2 3		s s s	55C2 55C3 55B5	decay scheme confirmed by (.123γ)(.014γ)(55M5, 55L5, 55M6).
				noβ^+ (55M5, 55C2, 55L5)						
				very weak β^+ (impurity?)(55M6, 55C3)						
Ni $\xrightarrow{57}$ Co 28 29 27 30	3.25	1	36^h	β^+	0.854	9	76	s	56K3	(0.72β^+ + 0.127γ + 1.37γ) confirm Q. (56K3)
$3/2^-$ 5.3 $5/2^-$, $3/2^-$ $7/2^-$				γ	1.363 1.39	8 2*		scin scin	56K3 54M5	
Co $\xrightarrow{58}$ Fe 27 31 26 32	2.319	6	72^d	β^+	0.485	10		s	55C3	no 0.5γ (56G3, 56F1)
2^+ 2^+				γ	0.805 0.815	8* 4*		scin scin	56G3 55C3	
6.6 2^+ 0^+				also weak ϵ branch (56G3, 56F1)						
Cu $\xrightarrow{52}$ Ni 29 30 28 31 $3/2^-$ $7/2^-$ $3/2^-$	3.7	3	82^s	β_o^+	3.75 3.4	10 5		s a	56S8 55L6	β_o^+ is ~ what is expected from β-decay systematics (54W2). However (55Y1) reports 1.85 β^+ accompanied by only a weak γ. Weak γ's also reported by (56P4).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Co ⁶⁰ ---> Ni 27 33 28 32 5^+ 7.4 4^+ 12.6 2^+ 0^+	3.818	3	5.2^y	β^-	0.309	3		s	54B2	1.48 β^- (5^+ ---> 2^+) <u>can</u> be fit with ΔJ = 3; no shape (56W1)
Cu ⁶⁰ ---> Ni 29 31 28 32 2^+ 2^+ 2^+ 0^+	6.27	3	24^m	β^+	3.92	3		s	54N4	
Zn ⁶⁰ ---> Cu 30 30 29 31 0^+ 2^+	>0		2.1^m	weak γ with $0.5<E_\gamma<3$					56L3	Produced by 52 mev α on Ni 58 but not by 15 mev α. β systematics (54W2) suggest Q ~ 4
Fe ⁶¹ ---> Co 26 35 27 34 5/2, $3/2^-$ $7/2^-$	> 0		5.5^m	β^-, γ					55R4a	β systematics (54W2) suggest Q ~ 3
Co ⁶¹ ---> Ni 27 34 28 33 $7/2^-$ 5.2 $5/2^-$ $3/2^-$	1.29	4	1.66^h	β^- γ	1.22 1.3 1.1 .069 .066	4 1* 1* 2* 2*	95	scin a a scin scin	56N2 55F1 55N2 56F2 54N5	no 0.5γ (56N2)
Cu ⁶¹ ---> Ni 29 22 28 23 <u>$3/2^-$</u> 5.2 $3/2^-$	2.24	1	3.3^h	β^+_0	1.220	15	50	s	56N2	energy match by ($0.56\ \beta^+$)(0.66γ)(56N2)
Zn ⁶¹ ---> Cu 30 31 29 22 $3/2^-$ 5.4 <u>$3/2^-$</u>	5.8	3	1.5^m	β^+_0	4.9 4.8	5 3		scin a	55C4 55L3	no strong γ (55C4)
Co ⁶² ---> Ni 27 35 28 34 ? ≥5.8 2^+ 0^+	≥3.8	2	14^m	β^- γ	2.8 1.17	2 3*	 str	a scin	54N6 54N6	
Cu ⁶² ---> Ni 29 33 28 34 0^+, 1^+ ≥5.1 0^+	3.03	1	9.7^m	β^+_0	2.91	1		s	54N6	no γ (55R3)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Zn -62-> Cu 30 32 29 33 0^+ ≥4.7 1^+, 0^+ 0^+, 1^+	1.69	5	9.3^h	β^+ γ	0.675 0.0413	10 3		s s	54N6 54N6	β^+ may be 2 spectra of approx. = intensity diff. in end point by ~.04.
Zn -65-> Cu 30 35 29 36 $5/2^-$ 8.4 $3/2^-$	1.347	1	245^d	β^+	0.325 0.324	2 4		s s	56A2 56A3	
Ni -66-> Cu 28 38 29 37 0^+ 5.7 1^+	0.20	3	55^h	β^-	0.20 ~0.15 ~0.3	3		scin a a	56J5 56K4 55N2	F-K plot linear (56J5); no γ (56J5)
Cu -66-> Zn 29 37 30 36 1^+ 2^+ 5.4 0^+	2.63	1	5.1^m	β_0^-	2.63	3		scin	56J5	Energy match provided by 1.64β + 1.03γ.
Ga -67-> Zn 31 36 30 37 $3/2^-$ $5/2^-$	≥0.88 ≤ 1.1	1	78^h	γ	0.870	10		scin	55R4	Complex γ's consistent decay scheme with highest level in Zn^{67} at 0.87. Upper limit on Q from no β^+ and ft value for ~ 1 per cent ϵ branch to 0.87 state.
Ge -67-> Ga 32 35 31 36 $3/2^-$, $5/2^-$ ≥5.3 $3/2^-$	≥ 3.9	3	19^m	β^+	2.9	3		a	56A4	Recalibration of absorbers with known radiations leads to reinterpretation of previously reported 3.4β^+ as 2.9 β^+. (56A4)
Ga -68-> Zn 31 37 30 38 1^+ 2^+ 0^+	2.91	2	68^m	β_0^+	1.94	5		s	56C2	
Ge -68-> Ga 32 36 31 37 0^+ 1^+	≥ 0 ≤ 1.3		275^d	no β^+ with E > 0.3				s	56C2	β-decay energy systematics (54W2) suggest Q ~ 0.7.
Zn -69-> Ga 30 39 31 38 $1/2^-$ 4.4 $3/2^-$	0.905	6	52^m	β_0^-	0.85 0.914	4* 5*	100	s	55L7 53S1	simple β^- decay (53S1); no $52^m\gamma$ (55L7)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
As $\xrightarrow{69}$ Ge 33 36 32 37 $3/2, 5/2^-$ ≥ 5.5 $5/2, 3/2^-$	3.9	3	15^m	β^+	2.9	2*		a	55B6	from J, π assignments to g.s. of As^{69}, Ge^{69}, J seems probable that β^+ is a composite spectrum.
				γ	0.23	1*		scin	55B6	
Ga $\xrightarrow{70}$ Ge 31 39 32 38 1^+ 5.1 0^+	1.65	1	21.1^m	β_0^-	1.65	1	99.2	scin	57B2	
As $\xrightarrow{70}$ Ge 33 37 32 38 $4,5^+$ 6.2 0^+	6.54	10	52^m	β_1^+	2.2 2.45	3 4	 33	a s	56A4 55S4	a consideration of states from complex decay of Ga^{70} and additional γ's found by (55B6) support decay scheme and adopted Q.
				β_2^+	1.35	3	67	s	55S4	
Ge $\xrightarrow{71}$ Ga 32 39 31 40 $1/2^-$ 4.3 $\underline{3/2^-}$	0.233	3	12^d	continuous γ-spectrum Eγ max						K binding energy is 0.0111
					0.226	5		scin	56L4	
					0.220	3		scin	55B7	
As $\xrightarrow{71}$ Ge 33 38 32 39 $5/2^-$ 5.7 $3/2^-, 5/2^-$ $1/2^-$	2.011	8	62^h	β^+	0.815 0.813	10 10	 97	s s	55G2 54T1	$(0.8\beta^+)(0.175\gamma)$ observed by 54T1
				γ	0.1750 0.1745	4 4		s s	55G2 54T1	
Ga $\xrightarrow{72}$ Ge 31 41 32 40 $3, 4^-$ 9.0 2^+ 0^+	4.00	2	14.1^h	β^-	3.166	30*	8	s	55J1	Very complex β and γ spectra, but consistent decay scheme and Q are established by βγ and γγ (56K5). [eg. (2.53β)(0.63γ)(0.83γ)]
				γ	0.834	5*		scin	55J1	
				γγ, βγ					56K5	
Se $\xrightarrow{72}$ As 34 38 33 39 0^+ 2^-	≥ .045		8.4^d	γ	0.0451			xγscin	57J2	β-decay energy systematics (54W2) suggest Q ~ 0.2
Se $\xrightarrow{73}$ As 34 39 33 40 $9/2^+$ 5.4 $9/2^+$ $5/2^-$ $3/2^-$	2.75	1	7.1^h	β^+	1.29	1	100	s	56H3	[≯ 1 per cent 1.65 β (56H3, 56A4)] and (γ±)(0.066γ (0.36γ), (56H3) establish decay scheme and Q.
				γ	0.0658	1		s	56H3	
				γ	0.359	1		s	56H3	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Ga -74-> Ge 31 43 32 42 0^+	≥ 2.6	1	~ 8^m	γ	0.58, 2.3, 2.6			scin	56M2	β-decay energy systematics (54W2) suggest Q ~ 6.
Ge -75-> As 32 43 33 42 $1/2^-$ 5.1 $3/2^-$	1.150	15	82^m	β_0^-	1.188	20	87	scin	55S5	(11 per cent 0.919β)(0.265γ) confirm Q. (55S5)
Se -75-> As 34 31 33 42 $5/2^+$ $3/2^-$	> 0.4		127^d	no ε to g. s.					55S5	decay scheme may be established by γγ (55S5) and by Σ scin (55L8) with highest state at 0.4.
As -76-> Se 33 43 34 42 2^- 8.2 2^+ 8.8^1 0^+	2.97	1	26.5^h	β_0^-	2.97 2.965	1 10	 50	s s	56P1 55K2	complex decay. Q confirmed by (2.42β)(0.55γ). (56P1, 55K2). ΔJ = 2; yes shape for β_0(56P1, 55K2)
Kr -76-> Br 36 40 35 41 0^+ 2^-	≥ 0.4		10^h	no γ± γ 0.4				scin scin	55T2 55T2	other γ's observed (55T2). β decay energy systematics (54W2) suggest Q ~ 1
Ge -77-> As 32 45 33 44 $1/2^-(52^s)$ $7/2^+$ $3/2^-$	2.74	6	12^h	data from 52^s isomer decay β_0^- β_1^- γ_1 γ_2	 2.90 ~2.7 0.159 0.215	 5 3 3	 90 10 	 scin aβγ scin scin	 55B4 54B3 54B3 54B3	(2.7β)(0.215γ); no(0.159γ)(0.215γ); no β(0.159γ). (54B3). 0.159 is excitation energy of 52^s state above g. s. of Ge^{77}
Kr -77-> Br 36 41 35 42 $1/2^-$ 5.4 $3/2^-$	2.88	2	1.2^h	β_0^+	1.86	2*	61	s	55T2	complex decay. lower energy β^+ groups and γ's present.
Kr -79-> Br 36 43 35 44 $1/2^-$ 5.5 $3/2^-$	1.619	5	34.5^h	β_0^+ no (0.598β^+)γ	0.598	5		s	54T2 54T2	F-K plot is linear for β_0.
Br -80-> Kr 35 45 36 44 1^+ 5.5 0^+	2.00	1	4.5^h	β_0^-	1.99	1	85	s	56L5	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Br -82-> Kr 35 47 36 46 5^- 5.1 $4,5^-$ 0^+	3.07	1	35.5^h	β^-	0.460 0.444 0.435 0.440	10 1 5 10		scin s s s	56D2 56W2 56B2 53L1	complex γ cascades. Several investigators give similar decay schemes.
				β^- feeds state which decays by γ cascades. state energy	2.620 2.615* 2.645* 2.620	10* 5* 7* 10		$\gamma\gamma$scin $\gamma\gamma$scin $\gamma\gamma$scin Σ scin	56D2 56W2 56B2 54L2	
Rb -83-> Kr 37 46 36 47 $5/2^-$ $3/2^-$ $9/2^+$	> 0.57		~ 90^d	γ (K Xray) (0.525γ)	0.525	7		scin	55P2 55P2	no 0.8γ or 0.15γ. γ decay is to 1.9^h Kr^{83}
Br -84-> Kr 35 49 36 48 2^-, 1^- 7.6 0^+	4.70	5	33^m	β_0^-	4.76	10*	25	scin	57J3	complex decay (3.83β) (0.879γ) confirm Q.
Rb -84-> Kr 37 47 36 48 2^- 8.5^1 0^+	2.66	2	33.0^d	β_0^+	1.70 1.66	10* 2		scin s	55W3 56W4	$(\gamma\pm)(\gamma)/\gamma$ indicate β_0 is g. s. transition (55W3)
Rb -84-> Sr 37 47 38 46 2^- 9.2^1 0^+	0.91	1	33.0^d	β_0^-	0.91	1	~3	s	56W4	
36 Kr 49 -85-> Rb 37 88 $9/2^+$ 8.2^1 $5/2^-$	0.675	7	10.6^y	β_0^-	0.672	7	>99	s	55T2	0.672β^- shows ΔJ = 2, yes shape(55T2). 0.824β + 0.1495γ − 0.305γ from 4.4^h isomeric state gives energy match.
Rb -86-> Sr 37 49 38 48 2^- 8.4^1 0^+	1.772	5	18.65^d	β_0^-	1.795 1.760 1.770 1.767 1.798 1.770 1.760	15 10 10* 15* 15* 15* 15*	86 85 85 80 88 	s s s s s s s	56A2 56L5 55L9 55K5 54C1 54P3 54M6	ΔJ = 2, yes shape observed for β_0^- (56A2, 56L5, 56L9, 55K5, 54C1, 54P3, 54M6). s $\beta\gamma$ gives linear F-K Plot and energy match. (1.08γ) (0.71β). (56L5, 55L9, 54C1, 54P3)
Y -86-> Sr 39 47 38 48 ? 6.7 0^+	6.01	5	15^h	γ_1 γ_2 γ_3 γ_4	1.93 1.08 0.635 0.180			scin	54H1	Each γ coinc. with all others and with $\gamma\pm$ by Σ scin (54H1). Thus 1.8β^+ + 1.2β^+(51 H 24) feed parent states in cascade.

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Zr -86-> Y 40 46 39 47 0+ ?	≥ 0.24	3	17^h	ε						
				K x ray				scin	59H1	
				γ	0.241	7*		scin	54H1	
				no β^+				scin	54H1	
Kr -87-> Rb 36 51 37 50 7/2+ 7.5 5/2− 3/2−	4.2	2	78^m	β_1^-	~3.8		≥65	sβγ	55T2	(3.8β)(0.403γ), (3.3β, 0.85γ), and (1.3β)(2.57γ)(.403γ) provide energy match. (55T2)
				γ_1	0.403	4		scin	55T2	
				βγ					55T2	
Rb -87-> Sr 37 50 38 49 3/2− 17.6 9/2+	0.273	4	~5x10^{5}y	β^-	0.250	25		a	54F3	3rd forbidden transition
Y -88-> Sr 39 49 38 50 4− 3− ≥9.0 2+ 0+	3.45	4	105^d	β^+	0.58	1*	<0.3	s	56S9	$0.83\beta^+$ previously assigned, due to internal pairs from 1.85γ, (56S9)
				γ	1.85	4*		scin	56L6	
					1.85	4*		scin	56S10	
					1.83	4*		scin	55A1	
Rb -89-> Sr 37 52 38 51 3/2− 7.8 5/2+	3.90	10	15^m	β_o^-	3.92	5	7	scin	5602	complex decay 2.81β(5 per cent) + 1.05γ give energy match
Sr -89-> Y 38 51 39 50 5/2+ 8.3^1 1/2−	1.463	10	51^d	β_o^-	1.462	5	100	s	55B8	β_o shows ΔJ = 2;yes shape (55B8)
Nb -89-> Zr 41 48 40 49 9/2+ 6.1 9/2+	3.88	10	1.9^h	β_o^+	2.85	10		scin	55M7	
					2.9	4		a	54D2	
				noγ				scin	55M7	
Y -90-> Zr 39 51 40 50 2− 0+	2.26	1	65^h	β_o^-	2.275	10		s	56P3	β_o shows ΔJ = 2;yes shape (56P3, 55J2)
					2.26	2		s	55J2	
Nb -90-> Zr 41 49 40 50 ? 0+	~6.2		15^h	β^+	1.51	3		scin	56L7	A decay scheme can be set up similar to that of (55T3) incorporating the data of (56L7) but leading to a Q value ~6.2 rather than 6.0 as proposed by (55T3)
					1.7	1		scin	55M8	
					1.500	5	87	s	54O1	
				complex γ spectrum. Various data and decay schemes presented. (56L7, 55M8, 54O1):						

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Mo -90-> Nb 42 48 41 49 0+ 5.2 0+, 1+ ?	2.54	10	5.7h	β+	1.15	10		scin	55M8	5.7h Mo90 ----> 0.015s Nb90 ----> 24s Nb90 ----> 15h Nb90 (g. s.)
				γ1	0.250	2*		s	55M8	
				γ2	0.120	1*		s	55M8	
Sr --91-> Y 38 53 39 52 5/2+ 8.2[1] 1/2-	2.67	1	9.7h	$β_0^-$	2.7	3*	25	a	55H3	
Y -91-> Zr 39 52 40 51 1/2- 8.8 1/2, 3/2+ 8.5[1] 5/2+	1.549	4	57.5d	β-	0.36 0.33	2* 1*		βγa βγscin	55K6 54B4	confirms and assigns previously reported weak γ. Good energy match to g. s. transition.
				γ	1.190 1.22	5* 1	0.22 0.3	scin scin	55K6 54B4	
Nb -91-> Zr 41 50 40 51 1/2- 60d 9/2+ 1/2, 3/2+ 5/2+	1.1 ≤ Q ≤ 1.7		?	ε see comment						ε observed is from 60d metastable state. Q > 1.21 - 0.104. Absence of β+ and consideration of the ΔJ = 2; no nature of transition sets upper limit on Q (55T4)
				Zr x ray				crit abs	55H4	
				No β+				scin	55H4	
				γ1	0.104	1*		s	55H4	
				γ2	1.21	1*		scin	55H4	
Mo -91-> Zr 42 49 41 50 9/2+ 5.6 9/2+	4.42	4	15.7m see comment	$β_0^+$	3.44 3.32 3.2	3 5 3*	94	s a a	56S11 53K1 49D1	Consideration of (γ, n) thresholds for 66s and 16m Mo91, the centrifugal barrier operative in the case of the 9/2+ state and the existence of a 0.65γ in the 66s state assigns 16m Mo91 as the 9/2+ ground state. (55A2).
				no γ					56S11, 55A2, 49D1	
Sr -92-> Y 38 54 39 53 0+ 4.4 1+ 2-	1.92	4	2.7h	β-	0.545	5	90	scin βγ	57H1	weak 0.44γ, 0.23γ (57H1)
				γ	1.37	5		scin	57H1	
Y -92-> Zr 39 53 40 52 2- 8.2[1] 0+	3.60	5	3.5h	$β_0^-$	3.6 3.60	2 5*	 77	s s	55C5 54A1	ΔJ = 2; yes shape observed (54A1). γ present (55C5, 54A1). Energy match provided by (1.3β)(1.4γ) + 0.93γ (55C5)
Nb -92-> Zr 41 51 40 52 2,3 2+ 0+	1.84 ≤ Q		10d	ε						(γ1) (γ2), (γ2, γ3) (K x ray) (55H4)
				γ1	0.900	10*		scin	55H4	
				γ2	0.934	10*		scin	55H4	
				γ3	1.83	2*		scin	55H4	β+ < 10^-4 (55H4)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Tc -93-> Mo	3.16	3	2.7^h	β_1^+	0.82	1*		s	54L3	(β_1^+) (γ_1) and (β_2^+) (γ_2), (54L3).
43 50 42 51				β_2^+	0.64	2*		s	54L3	Relative intensities of groups
$9/2^+$ -4.9 $7/2^+$				γ_1	1.35	2*		scin	54L3	indicate β^+ transitions are allowed,
$5/2^+$				γ_2	1.50	5*		scin	54L3	hence $9/2^+$ for Tc^{93}.
Nb -94-> Mo	2.13	5	$2 \times 10^4 y$	β^-	0.5-0.6			a	55R5	1.65 γ cross over (55R5)
41 53 42 52				γ_1	0.726	10*		scin	55R5	
6^+ 12.1 4^+				γ_2	0.903	10*		scin	55R5	
2^+										
0^+										
Zr -95-> Nb	1.122	5	65^d	β_o^-	1.13	4		s	56A2	0.364β + 0.756γ and 0.9β to 0.235
40 55 41 54										
$5/2^+$ 6.9 $5/2, 7/2^+$				β_1^-	0.404	8		s	56A2	Nb^{95} state provide energy matches
				γ_1	0.723 0.722	2 3 *		s s	56A2 55R6	(56A2)
>10.2 $9/2^+$										
Nb -95-> Mo	0.929	4	35^d	β^-	0.162	5		s	56A2	g. s. J of Mo^{95} has also been reported
41 54 42 53										
$9/2^+$ 5.1 $7/2, 9/2^+$				γ	0.764 0.767	2 2		s scin	56A2 56J3	as 7/2 by (55M9). 5/2 seems more
$5/2^+$ see comment										consistent with the absence of a g. s.
										transition.
Ru -95-> Tc	2.2 or 2.6	1 1	1.6^h	β^+	1.2	1		a	56S12	observed ϵ/β^+ and γ intensities
44 51 43 52										suggest β^+ is either to g. s. or to
$5/2^+$, $7/2^+$ >4.8 $3/2^+$ $7/2^+$										level at 0.38. (0.34γ + 0.039 isomeric
$9/2^+$										decay.)
Zr -97-> Nb	2.65 see comment	2	17^h	β^-	1.900	25		s	56S13	Shell model favors $5/2^+$ or $7/2^+$ for
40 57 41 56										
? 60^s, $1/2^-$				γ	0.746 0.75 0.75	5 1 1		s scin scin	56S13 56N3 55S6	Zr^{97}. Data do not rule out pos-
see comment 7.1 $9/2^+$										sibility of ΔJ = 2, yes β transition
										(i. e. $5/2^+$ assignment). In this case,
										log f^1t = 8.5 and Q ~ 2.6.
Nb -97-> Mo	1.90	4	74^m	β^-	1.180	25		s	56S13	g. s. J of Mo^{97} has also been report-
41 56 42 55										
$9/2^+$ 5.4 $9/2^+$				γ	0.664 0.66 0.67	6 1 1*		s scin scin	56S13 56N3 55S6	ed as 7/2 by (55M9). For this spin,
										a 1.8β would exist, but might be
$5/2^+$ see comment										masked by 1.9β of Zr^{97}

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
$Ru_{44\ 53} \xrightarrow{97} Tc_{43\ 54}$ $5/2^+$ 91^d $1/2^-$ $9/2^+$	≥ 0.41		2.44^d	γ	0.219 0.220 0.2161	4 5* 10*		s scin s	56S13 55A3 55C6	(0.109γ)(.216γ) and x(0.109γ, 0.325γ) indicate level 0.325 above 91^d Tc^{97}. (55C6). Very weak 0.57γ also seen (55C6).
$Tc_{43\ 55} \xrightarrow{98} Ru_{44\ 54}$ 6^+ ~12 4^+ 2^+ 0^+	1.70	5	$\sim 10^{5y}$	β^-	0.30	5*		scin	55K7	(0.3β)(0.65γ, 0.74γ). (55K7). (0.66γ)(0.75γ). (56B3). no x (0.65γ, 0.74γ) (55K7).
				γ_1	0.660 0.65	10* 1*		scin scin	56B3 55K7	
				γ_2	0.750 0.74	10* 10*		scin scin	56B3 55K7	
$Rh_{45\ 53} \xrightarrow{98} Ru_{44\ 54}$ 2, 3^+ ≥ 4.7 2^+ 0^+	4.35	15	8.7^m	β_1^+ β_2^+	2.5 3.3	2 4		scin a	56K6 55A3	no other γ (56K6). Other γ's with Eγ > 0.7 (55A3)
				γ	0.650 0.650	10 20		scin scin	56K6 55A3	
				$(2.5\beta^+)$ (0.65γ)				scin	56K6	
$Pd_{46\ 52} \xrightarrow{98} Rh_{45\ 53}$ 0^+ 2, 3^+	> 0		17^m					scin	56K6 55A3	
$Mo_{42\ 57} \xrightarrow{99} Tc_{43\ 56}$ $1/2^+$ 7.1 $1/2^-$ $9/2^+$	1.37	1	68^h	β_1^-	1.18	2*		s	54L4	$(\beta_2)(\gamma_2)$ by (54L4, 54V1) provides energy match.
				β_2^-	0.80 0.87	2* 4*		sβγ scin βγ	54L4 54V1	
				γ_2	0.390 0.372	20 10		sγβ scin	54L4 54V1	
$Rh_{45\ 54} \xrightarrow{99} Ru_{44\ 55}$ $7/2^+$, $9/2^+$ 4.9 <u>$5/2^+$</u>	2.08 see comment	3	4.7^h	β^+	0.74	2*	10	scin	56K6	15^d Rh^{99}(?) with (0.35γ)(0.09γ) reported by (56H4, 55F2). Possibly isomeric with 4.7^h Rh^{99}.
				γ	0.355	10*		scin	56K6	
				$(0.74\beta^+)(0.355\gamma)$				scinβγ	56K6	
$Pd_{46\ 53} \xrightarrow{99} Rh_{45\ 54}$? >4.8 7/2, $9/2^+$	3.8	1	22^m	β^+	2.0	1		scin, a	56K6	other β^+, other γ present. p. 4.7^h Rh^{97}, β decay energy systematics (54W2) suggests Q ~ 3.
				γ_1	0.670	10*		scin	56K6	
				γ_2	0.140 0.132	5* 5*		scin scin	56K6 55A3	
				$(2.0\beta^+)(0.140\gamma, 0.67\gamma)$ $(0.14\gamma)(0.67\gamma)$				γγ scin βγscin	56K6 56K6	
$Mo_{42\ 59} \xrightarrow{101} Tc_{43\ 58}$ $5/2^+$ 6.0 10^{-3s} 3/2, $5/2^-$ $9/2^+$	2.81	5	14.6^m	β_1^-	2.23 2.20	4* 6*		scin βγscin	5701 56M4	Complex β and γ spectra. (5701, 56M4). (56M4) suggests a level scheme. $t_{1/2}$ (0.19 state in Tc^{101}) ~ 10^{-3s}. (56M4) 0.19γ is M2 or E3.
				γ_1	0.590 0.585	10* 10*		scin? scin	56M4 5701	
				(β_1, γ_1)				scin	56M4	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Tc -101-> Ru	1.63	3	14^m	β_1^-	1.32	3	strong	scin	5701	
43 58 44 57				β_2^-	1.07	5		scin	5701	
$9/2^+$ 4.5 $7/2^+$				γ_1	0.307	10*		scin	5701	
$5/2^+$				γ_2	0.545	10*		scin	5701	
					0.540	8		scin	56T2	
				$\beta_1\gamma_1$, $\beta_2\gamma_2$ observed					5701	
Rh -101-> Ru	> 0.31		see comment	ϵ (> 3y Rh^{101})						identification of Rh^{101} is (?)(56H4)
45 56 44 57				γ_1	0.190	5*		scin	56H4	but γ's fit levels seen in $Ru^{101}(\alpha,\alpha')$
$9/2^+$ (?)				γ_2	0.125	5*		scin	56H4	(55T5). It may be that these two
$5/2^+$				$\gamma_1\gamma_2$ observed				56H4,	56P5	activities are closely spaced
				ϵ	4.7^d Rh^{101}					isomers.
				γ_3	0.312	10			56K6	
				no other γ					56K6	
Pd -101-> Rh	1.57	4	8.5^h	β_0^+	0.58	4	~ 4	scin,	56K6	complex γ spectrum (56K6)
46 55 45 56				no β^+ γ				scin	56K6	
$7/2^+$(?) ≥ 5.2										
$9/2^+$(?)										
Rh -102-> Ru	2.28	2	215^d	β_0^+	1.27	2*		s	54K3	β^+ spectrum complex (54K3) (54M7)
45 57 44 58					1.240	20*	75+	s	54M7	Mixed β^+, β^- present. Reasonable
2^- ? 8.1[1]				complex γ's				56A3,	56H4	energy match by $0.76\beta^+ + 0.474\gamma$,
0^+								55D3, 55B9,	55F2	(54M7). All γ are (x)(γ), (55D3).
								54K3,	54M7	
Rh -102-> Pd	1.13	2	215^d	β_0^-	1.12	2*		s	54K3	Mixed β^+, β^-, present. β^- spectrum
45 57 46 56					1.150	20		s	54M7	complex (54K3), F-K plot linear (54M7)
2^-? 8.8[1]										Presence of e^- makes decision on
0^+										complexity of β^- weak. However,
										(55D3) reports x(all γ), hence pro-
										bably all γ in β^+ or ϵ branch.
Ru -103-> Rh	0.748	5	39.5^d	β^-	0.216	8		s	56S13	(.12β) (0.61γ), (55S7, 55D3) pro-
44 59 45 58					0.202	2		s	55F3	vides good energy match.
? 5.7 $3/2^+$				γ	0.496	4		s	56S13	
56^m $7/2^+$					0.498	5*		scin	55F3	
$1/2^-$					0.495	5*		scin	55S7	
					0.498	4		s	54D3	
				From 56^m Rh^{103} isomer						
				γ	0.038	2		s	56S13	
					0.0402	5		s	55A4	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Pd -103-> Rh 46 57 45 58 5/2, 7/2$^+$ ≥ 5.7,56^m 7/2$^+$ 1/2$^-$	0.56	2	17^d	ε γ	 0.495 0.498	 5 4		 scin scin	 55S7 55A4	K binding energy = 0.023 ~ 0.5 state fed by L capture only. (55A4)
Ag -103-> Pd 47 56 46 57 9/2$^+$ ≥ 4.3, 7/2$^+$, 5/2$^+$	2.3 or 2.9	2 2	59^m	β$^+$ γ	1.3 0.554 0.764	2 5* 8*		a s	54H2 53B1	First excited state of Pd is at ~ 0.5 excitation by Rh^{103}(p,n). (54P4).
Tc -104-> Ru 43 61 44 60 ? ≥ 5.2 0$^+$	~4		3.8^m	β$^-$ γ	~2 >1			a scin	56F3 56F3	Eγ/β ~ indicates Q ~ 4 (56F3) 18^m Tc^{104} has Q ~ 5.5, so 3.8^m is probably g. s. (56F3).
Rh -104-> Pd 45 59 46 58 1$^+$ 2$^+$ 4.5 0$^+$	2.44	3	44^s	β_0^-	2.44	3*		s	55B10	energy match by (1.88β$^-$) (0.556γ) and (0.64β$^-$) (1.24γ), (55B10).
Ag -104-> Pd 47 57 46 58 3$^+$ (?) 5.3 2$^+$ 0$^+$	4.28	1	27^m	β$^+$ γ	2.70 0.5562	1 4		s s	55J3 55J3	β$^+$ and γ probably in cascade. F-K plot of β$^+$ straight above endpoint of Cd^{105} β$^+$ contamination (1.7 mev), (55J3).
Cd -104-> Ag 48 68 47 57 ≥ 0.13 3$^+$(?)	≥0.13		59^m	ε no β$^+$ γ	 0.1342	 1		 s s	 55J3 55J3	p. 27^m Ag^{104}. Other lower energy γ assigned to decay present as well as higher energy unassigned γ's, (55J3).
Ru -105-> Rh 44 61 45 60 ? 5.8 ? 1/2$^-$ 7/2$^+$	2.008	7	4.5^h	β$^-$ γ from 45^s Rh^{105} isomer γ	1.15 0.728 0.133	2 4 10		s s s	56S13 56S13 56S13	
$_{45}Rh_{60}$ -105-> $_{46}Pd_{59}$ 7/2$^+$ 5.6 5/2$^+$	0.565	3	36.5^h	β_0^-	0.560 0.560	15 5	 70	s s	56S13 56L5	0.25β + 0.31γ gives good energy match, (56S13). (0.25β)(0.31γ), (56L5).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Ag $\xrightarrow{105}$ Pd 47 58 46 59 1/2 ? 5/2⁺	≥ 1.1		45ᵈ	ε						
				γ_1	0.443 0.446 0.433 0.430	1 4 * 7 9 *	30⁺	s s scin scin	56H5 55A5 55T6 55M10	γ intensities relative to 0.345γ = 100.
				γ_2	0.654	5	8⁺	scin	55H5	
				$\gamma_1\gamma_2$				scin	55H5	
Rh $\xrightarrow{106}$ Pd 45 61 46 60 1⁺ 5.2 0⁺	3.54	2	30ˢ	β_0^-	3.58	4		s	56S14	
Ru $\xrightarrow{107}$ Rh 44 63 45 62 ~ 6.2	≥ 4		4ᵐ	β^-	~4			a	55B11	high energy γ present (55B11)
Rh $\xrightarrow{107}$ Pd 45 62 46 61 ? ≥ 4.9 5/2⁺	> 2 or > 1 see comment		see comment	24ᵐ	Rh^{107}					possible Rh^{107} activity $t_{1/2}$ = 2.3^h, β^- ~ 1 and complex γ reported. (55N2). It is more likely that this is from an isomer of Rh^{107} than from Rh^{106}.
				β_1^-	1.15	10*	~100	a	55B11	
				β_2^-	~2		weak	a	55B11	
				complex γ spectrum				scin	55B11 55N2	
In $\xrightarrow{107}$ Cd 49 58 48 59 9/2⁺ 5/2, 7/2⁺ ≥ 5.0 7/2, 5/2⁺	~ 3 or ~ 3.2 see comment		33ᵐ	β^+ ~2				scin	55C7	0.22γ shown to be in coincidence with β^+(energy range 0.8-1 mev.) (55C7).
				γ	0.22	1		scin βγ, scin	55C7	
Rh $\xrightarrow{108?}$ Pd 45 63 46 62 ≥ 5.3 0⁺	≥ 4		18ˢ	β^-	4 - 5			a	55B11	not p. 13h Pd^{109} (55B11). Q too high to be any other Pd isotope (54W2)
				complex γ spectrum				a, scin	55B11	
Ag $\xrightarrow{108}$ Cd 47 61 48 60 1⁺ 4.6 0⁺	1.75	5	2.3ᵐ	β_0^-	1.7	1		a	56J6	0.639γ present (56J6)
Cd $\xrightarrow{109}$ Ag 48 61 47 62 5/2, 7/2⁺ ~ 4.8 39ˢ 7/2⁺ 1/2⁻	0.16		350ᵈ	ε						
				E_ϵ	~ .067 from ϵ_L/ϵ_K		≈0.32	4		
								pc, scin	54B5	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
$_{50}Sn_{59} \xrightarrow{109} {}_{49}In_{60}$	≥ 1.1		18^m	β^+				s	55P3	Complex γ (56P6). g. s. spins assigned by (56P6). 0.66γ is M4 (56P6). Beta decay energy systematics (54W2) suggest Q ~ 4.
$5/2^+$ $9/2^+$				γ_1	0.658	7*		s	56P6	
					0.678	7*		s	55P3	
				γ_2	1.12	10*		s	56P6	
$_{47}Ag_{63} \xrightarrow{110} {}_{48}Cd_{62}$	2.875	15	24^s	β^-_o	2.869	20	100	s	56T3	
1^+ 4.7 0^+					2.7	1		a	56J6	
					2.84	10		scin	54B6	
				no other γ present				scin	56T3	
				other γ present				scin	56J6 54B6	
				no $\beta_o\gamma$					56T3	
$_{50}Sn_{60} \xrightarrow{110} {}_{49}In_{61}$	> 0.28	1	4.1^h	γ	0.285	3*		s	55M11	p. 66^m In^{110}. β-decay energy systematics (54W2) suggests Q ~ 0.
0^+										
$_{46}Pd_{66} \xrightarrow{112} {}_{47}Ag_{65}$	0.30	2	21^h	β^-	0.28	2		aβγ	55N3	0.0185γ is <u>not</u> Ag x ray (55N3).
0^+ > 4.3 1^+				γ	0.0185	5		scin	55N3	
2, 1, 0^-				(0.28β)(0.0185γ) GM,				scin	55N3	
$_{48}Cd_{65} \xrightarrow{113} {}_{49}In_{64}$	0.28	3	≥ 3 x 10^{15y}	from 5.1^y Cd^{113}						
5^y $11/2^-$				γ	0.265	10*	0.1		56D3	~ 0.55 β^- from this state to In^{113} is known, thus establishing Q.
$1/2^+$ $9/2^+$				γ is E5					56D3	
$_{50}Sn_{63} \xrightarrow{113} {}_{49}In_{64}$	0.49	1	118^d	ε						
1/2, $3/2^-$				continuous γ spectrum						K binding energy = 0.028
4.7 1.7^h $1/2^-$				E_γ max	0.07	1*		scin	56J4	ϵ_L/ϵ_K = 0.17±0.12 (56A3)
$9/2^+$				x(continuum γ)				scin	56J4	
				From 1.7^h In^{113} isomer						
				γ	0.393	5		scin	56G4	
$_{49}In_{65} \xrightarrow{114} {}_{48}Cd_{66}$	1.42	3	72^s	β^+_o	0.400	25	.004	scin	56G5	previously reported 0.56γ and 0.72γ are associated with ε from 50^d In^{114}.
1^+ 5.1 0^+				1.3γ assigned to Sn^{114}					56J7, 56G5	
$_{49}In_{65} \xrightarrow{114} {}_{50}Sn_{64}$	1.985	4	72^s	β^-	0.674	40	0.09	scin βγ	56G5	
1^+ ~6 2^+					0.71	5	0.2	scin βγ	56J7	
4.5 0^+				γ	1.30	4	0.09	scin	56G5	
					1.299	2	0.2	s	56J7	
				(0.7β)(1.3γ)				scin βγ	56J7, 56G5	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Cd 115→ In 48 67 → 49 66 $1/2^+$ 7.1 4.5^h $1/2^-$ $9/2^+$	1.45	1	2.33^d	from 4.5^h In^{115}						
				γ	0.335	1*			55V2	
Sb 116→ Sn 51 65 → 50 66 3^+ 4^+ > 4.8 2^+ 0^+	4.73	9	15.5^m	β_1^+	2.3 2.4	3 3*		a scin βγ	56A4 55S8	
				β_2^+	1.5	2*		scin βγ	55S8	
				γ_1	1.305	15		scin	55S8	
				γ_2	2.2 2.215	1 20		scin scin	54A2 55S8	
				$(\beta_1)(\gamma_1)$, $(\beta_2)(\gamma_2)$					55S8	
In 117→ Sn 49 68 → 50 67 $9/2^+$ 4.6 $7/2^+$ $3/2^+$ $1/2^+$	1.462	10	1.1^h	β^-	0.74	1		s	55M12	
				γ_1	0.565	7		s	55M12	
				γ_2	0.161	1		s	55M12	
				also, from 1.9^h In^{117} isomer						
				β_o^-	1.772	5	55	s	55M12	
				γ_{IT}	0.311	2	22	s	55M12	
Sb 117→ Sn 51 66 → 50 67 $5/2^+$ $3/2^+$ $1/2^+$	1.70	15	2.8^h	γ	0.161	1		s	55M12	Q is calculated from β^+/ϵ for an allowed β^+ transition (55M12)
				β^+/ϵ = 0.026				scin (γ ±)γ	55M12	
Sb 118→ Sn 51 67 → 50 68 ? > 6 0^+	>1.7		5.1^h	β^+	0.7				53F1	3.3^m formerly considered g. s. of Sb^{118} is assumed to decay by 0.108γ I. T. To 5.1^h Sb^{118} (53F1). β-decay energy systematics (54W2) suggest Q ~ 3-4.
				$\beta^+\gamma$					53F1	
Sb 119→ Sn 51 68 → 50 69 $5/2^+$ 5.1 $3/2^+$ $1/2^+$	0.59	3	39^h	ε						K binding energy = 0.029; γ is M1(5604)
				continuous γ spectrum E_{max}	0.534	30		scin	5603	
				γ	0.02383		3	s	5604	
I 120→ Te 53 67 → 52 68 1,2^- >7 0^+	≥ 5		$>1.3^h$	β^+	4.0 4.0	2 2		s s	53F1 53D1	β-decay energy systematics (54W2) suggest Q ~ 5.

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Sb -122-> Sn 51 71 50 72 2^- 2^+ 0^+	1.55 see comment	5	2.75^d	β^+ ϵ γ	 1.152 1.137 1.10	 15 6 1*	~.01 <0.1 3 3.1 2.6 0.75 0.73	$\gamma\gamma$(e) s crit a pc pc s s scin	55G3 55F4 55G3 55F4 54J2 55G3 55F4 54C3	Q obtained using $\epsilon/\beta^+ = 300 \pm 120$, and $\Delta J = 2$,yes for β^+ transition. (55G3)
Sb -122-> Te 51 71 52 70 2^- 2^+ 2^+ 8.6^λ 0^+	1.973	5	2.75^d	β_0^-	1.970 1.987 1.97 2.00	5 20 1 2*	30 26 21 36	s s s s	55F4 55G3 54M8 54C3	excellent energy match by β + γ 1.400 10 + 0.566 4 55F4 1.423 10 + .563 5 55G3 1.41 1 + .588 5* 54M8 1.40 2* + .566 5* 54C3
I -122-> Te 53 69 52 70 1^+ 4.8 0^+	4.14	4	3.5^m	β_0^+ no γ	3.12	4		s scin	54M9 54M9	
Xe -122-> I 54 68 53 69 0^+ 1^+	> 0.23		19^h	ϵ , K x rays γ	 0.182 0.235			scin scin	54M9 54M9	β decay energy systematics (54W2) suggests Q ~ 0.3.
I -123-> Te 53 70 52 71 $5/2^+$ $3/2^+$ $1/2^+$	> 0.16		13^h	ϵ γ	 0.160	 2*		 s	 54M9	β decay energy systematics (54W2) suggests Q ~ 1.2.
Xe -123-> I 54 69 53 70 3/2, $1/2^+$ ≥ 5.1 $5/2^+$	2.7 or 2.9	1 1	1.8^h	β^+ γ	1.7 0.148	1* 2*		scin, a s	54M9 54M9	
Sb -124-> Te 51 73 52 72 4^+ 10.3 2^+ 0^+	2.925	15	54^d	β^- γ	2.300 2.39 2.31 0.605 0.605 0.603	10 3* 3* 6* 6* 10*	 22 22	s s s s s scin	56A2 55A6 54M10 56A2 55A6 54L5	non unique shape for 2.3 β^-, complex β and γ spectrum. Decay scheme providing good energy matches established.
Sn -125-> Sb 50 75 51 74 $3/2^+$ 9.5^m $11/2^-$ $3/2^+$ $7/2^+$ 9.0^1 $7/2^+$	2.35	2	9.7^d see comment	β_0^- γ $\beta\gamma$, $\gamma\gamma$	~ 2.4 ~ 2.4 1.41 1.41	 2* 2		scin a scin scin 56D2,	56D2 55B12 56D2 55B12 55B12	data of 56D2 and 55B12 fit a very consistent decay scheme (56D2). Lowest state to which $3/2^+$ 9.5^m Sn^{125} can decay is at 1.41 mev leading to Q ~ 3.4. Hence 9.7^d Sn^{125} is the ground state.

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
$Sb \xrightarrow{125} Te$ 51 74 52 73 $7/2^+$ 8.6^1 58^d $11/2^-$ $3/2^+$ $1/2^+$	0.757	7	2^y	β^-	0.612	12*	14	s	54M11	0.6β has ΔJ = 2, yes shape (54M11) Decay scheme with good energy matches from other β, γ present. Earlier data adjusted for shape to establish Q.
				βγ					54M11	
$Cs \xrightarrow{125} Xe$ 55 70 54 71 $5/2^+$? ≥ 5.1 $3/2^+$	3.07 or 3.18	2 2	45^m	β^+	2.05	2*		s	54M12	
				γ	0.112	1*		s	54M12	
				(γ±)(0.112γ)				s	54M12	
$I \xrightarrow{126} Te$ 53 73 52 74 2^- 2^+ 8.1^1 0^+	2.14	2	13.3^d	β^+_o	1.110 1.19	20 4*	0.96	s γ±γ scin	55K8 54P5	β^+_o has ΔJ = 2; yes shape (55K8). Good energy match by $0.46\beta^+$ + 0.65γ. Earlier data adjusted for shape to establish Q.
$I \xrightarrow{126} Xe$ 53 73 54 72 2^- 7.9 2^+ 8.8^1 0^+	1.235	15	13.3^d	β^-_o	1.250	10	5.8	s	55K8	β^-_o has ΔJ = 2; yes shape (55K8). Earlier data adjusted for shape to establish Q.
				β^-_1	0.865	5	29	βγs	55K8	
				γ_1	0.386 0.382	2 10*		s scin	55K8 54P5	
				(0.86β) (0.38γ)					55K8, 54P5	
$Cs \xrightarrow{126} Xe$ 55 71 54 72 1^+ ≥ 4.8 0^+	4.8	4	1.6^m	β^+	3.8	4		a,scin	54K4	absence of γ's from known transitions in Xe^{126}, and presence of 0.385γ is puzzling.
				γ	0.385	5		scin	54K4	
				no other γ (< 10 per cent)				scin	54K4	
				no β^+ (0.385)				scin	54K4	
$Ba \xrightarrow{126} Cs$ 56 70 55 71 0^+ 1^+?	> 0.7		96^m	no β^+ (?)				scin	54K4	0.9 ?γ and (0.7γ)(0.2γ) ?(54K4). β-decay energy systematics (54W2) suggest Q ~ 1.
				γ	0.700 0.225	30 10	33+ 100+	scin	54K4	
$Sb \xrightarrow{127} Te$ 51 76 52 74 $7/2^+$ 9.8^1 113^d $11/2^-$ $3/2^+$	1.64 see comment	2	93^h	β^-_o	1.57	3	30	scin	56D4	The allowed character of Te^{127} β^- decay considered together with the high ft value for β^-_o of Sb^{127} suggests that g. s. of Te^{127} is $3/2^+$ and β^-_o goes to $11/2^-$ 113^d Te^{127} isomer.
				β^-_1	1.11	3	20	βγscin	56D4	
				β^-_2	0.86	2	50	βγscin	56D4	
				γ_1	0.456	4		scin	56D4	
				γ_2	0.674	5		scin	56D4	
				$\beta_1\gamma_1$, $\beta_2\gamma_2$				scin	56D4	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Te ₋₋127₋> I 52 75 → 53 74 3/2⁺ 5.5 <u>5/2⁺</u>	0.690	7	9.4^h	β^-_o	0.695 0.72 0.683	10 5 10	99	s s scin	56K7 56K8 55D4	(56K7) reports weak (~ 1 per cent) γ spectrum
Xe ₋₋127₋> I 54 73 → 53 74 3/2⁺ <u>5/2⁺</u>	> 0.37 Q < 1.3		36.4^d	ε , no γ± γ	 0.368	 7*		scin scin	55M26 55M26	several other γ's present. γγ studies place highest level in I^{127} at 0.368. β decay energy systematics (54M26) suggest Q ~ 0.8. Upper limit set by absence of β^+.
Cs ₋₋127₋> Xe 55 7 → 54 7 5/2⁺ 5/2⁺ > 5.3 3/2⁺	2.09	2	6.25^h	β^+_o	1.063	11*		s	54M12	approximate energy match from 0.68β⁺ + 0.4γ (54M12).
I ₋₋128₋> Te 53 75 → 52 76 1⁺ 0⁺	> 0.75		25^m	γ ε/β+ε	0.750 =0.069	10*		xγscin pc,scin	56B4 56B4	$\varepsilon_{exc\,st}/\varepsilon_{g.s.}$ = 1/19 (56B4). β decay energy systematics suggest Q ~ 1.3.
I ₋₋128₋> Xe 53 75 → 54 74 1⁺ 6.6 0⁺	2.12	1	25^m	β^-_o	2.120	10	76	s	56B4	1.665β+ 0.455β give good energy match. (56B4)
Cs ₋₋128₋> I 55 73 → 54 74 1⁺ > 4.7 0⁺	4.0	1	3.8^m	β^+_o	3.0 ~2.9	1*	70	s a	55H5 55M13	lower β⁺ groups and γ's present (55H5).
Ba ₋₋128₋> Cs 56 72 → 55 73 0⁺ (?) 1⁺	> 0.27		2.4^d	ε γ	 0.270	 3*		 s	 55H5	β decay energy systematics (54W2) suggest Q > 0.
Sb ₋₋129₋> Te 51 78 → 52 77 7/2⁺ 7.4 ? 11/2⁻ 3/2⁺	≥ 1.98	5	4.6^h	β⁻	1.87	5	20	scin	56D4	several γ present, also lower energy β⁻. (56D4). β decay energy systematics (54W2) suggest Q ~ 3.
Te ₋₋129₋> I 52 77 → 53 76 3/2⁺ 5.9 5/2⁺ <u>7/2⁺</u>	1.481	5	74^m	β⁻ γ βγ, γγ	1.453 1.46 0.0268 ~0.035	5 1 5*	71 80	s scin s scin	56G6 55D4 56G6 55D4 55D4	decay scheme established with good energy matches.

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Cs -129-> Xe 55 74 54 75 7/2+ 1/2+	> 0.59		31^h	γ	0.55 0.585	3 20		scin scin	55R7 55N2	other γ's present (55R7, 55N2). β-decay energy systematics suggest Q ~ 1.4 (54W2).
Sb -130-> Te 51 79 52 78 6.2 0+	≥ 2.9	3	10^m	β⁻	2.9	2*		a	55F5	
Te -131-> I 52 79 53 78 3/2+ 6.2 5/2+ 7/2+	2.29	5	25^m	β⁻ γ 0.145γ/β ~ 1	2.14 0.145	5* 10	60	s scin GM	55H6 55M14 55M14	very complex decay makes energy determinations difficult. Decay scheme with fair energy matches (55H6).
Cs -131-> Xe 55 76 54 77 5/2+ 5.5 3/2+	0.354	5	10^d	ε continuous γ spectrum $E_γ$max	 0.320 0.320	 10 10*		 scin scin	 56H6 56M5	K binding energy = 0.0345
Ba -131-> Cs 56 75 55 76 1/2+ 5/2+	≥ 1.05	2	11.5^d	ε γ	 1.032 1.050 1.030 1.09	 15 15 15* 15*	 3+ 4+ (?)	 scin scin scin scin	 56B4 56G7 56V2 55L8	other γ's present, but γγ studies lead to decay scheme with highest level at ~ 1.0 (56B4, 56V2). 1.0γ intensities relative to 0.496γ = 100 β decay energy systematics (54W2) suggest Q ~ 1.7.
I -132-> Xe 53 79 54 78 4, 5⁻ 6.3 4+ 2+ 0+	3.56	3	2.4^h	β⁻ γ βγ, γγ	2.12 0.777 0.673	3* 8* 4*	18	s scin s	54F4 54F4 54F4 54F4	complex β and γ. Decay scheme consistent with coincidence and intensity data provides energy matches (54F4).
Cs -132-> Xe 55 77 54 76 2 2+ 0+	> 0.67		6.2^d	ε γ < 20 per cent ε to g. s. xγ/γ	 0.669			 scin	 55R8 55R8	weak 1.10γ, 1.26γ (55F8). However these γ's not seen in I^{132} decay. β decay energy systematics (54W2) suggest Q ~ 2.
Ba -133-> Cs 56 77 55 78 1/2+ 1/2+ 7/2+	0.48	4	10^y	ε $ε_L/ε_K$ to 0.44 level ≥ 9					 56L4	Intensity of K x ray due to K internal conversion only. (56L4) $ε_L/ε_K$ implies $E_ε$ ~0.04 (56L4)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF	COMMENTS
Cs -134-> Ba	2.055	3	2.3^y	β^-	0.650	5	56	s	55F6	extremely complicated decay. (4-5
55 79 — 56 78					0.655	2	50	s	55K9	β groups, ~ 10-12 γ rays) Disagree-
4^+ 8.9 4^+					0.645	10		s	55B13	ment on details of decay scheme. In
0^+				γ_1	0.604	1	weighted average of seven spectrometer measurements			general there is agreement that 0.65β
				γ_2	0.796	1				goes to state at ~ 1.4 mev.
				56J2, 55F6, 55B13, 55K9, 54V2, 54J3, 54B7.						
Pr -135-> Ce	≥ 3.5	1	22^m	β^+	2.5	1		a, scin	54H3	several γ (0.08 - 0.3) present,
59 76 — 58 77										(54H3)
$5/2^+$ > 5.1 $3/2^+$										
Cs -136-> Ba	2.85	1	12.9^d	β_1^-	0.657	3	7.4	s	54O2	complex γ spectrum. γγ and γ inten-
55 81 — 56 80				γ_1	1.041	5		scin,s	54O2	sities allow construction of con-
? 8.0				γ_2	0.822	2		scin, s	54O2	sistent decay scheme. Energy match
0^+				γ_3	0.335	2		scin, s	54O2	by 0.341β + 0.153γ + 0.162γ to 2.2
				γγ				scin	54O2	level. (54O2)
Pr -136-> Ce	≥ 3.0	1	70^m	β^+	2.0	1		a, scin	54H3	(0.17-1.1?) γ present (54H3)
59 77 — 58 78										
1^+ ≥ 5.2 0^+										
Xe -137-> Cs	≥ 3.5	3	3.9^m	β^-	3.5	3*		a	55N4	
54 83 — 55 82				γ					55N4	
$7/2^-$, $9/2^-$										
≥6.2 $7/2^+$										
La -137-> Ba	> 0.04		6×10^{4y}	ϵ, Ba K x ray				pc,crit a	56B5	
57 80 — 56 81										
$7/2^+$ >10 $3/2^+$										
Ce -137-> La	> 0.45		8.7^h	ϵ, La K x ray				pc,crit a	56B5 56D5	35^h Ce^{137} is $11/2^-$ isomer of 8.7^h
58 79 — 57 80	Q < 0.75			γ	0.445	5*	3	scin	56B5	$3/2^+$ Ce^{137} g. s. (56D5, 55B14).
$11/2^-$ 35^h					0.440	5*	2.3	scin	56D5	Upper limit on Q from ft value for
$3/2^+$ $1/2^+$										ϵ_K branches.
$5/2^+$										
$7/2^+$										
Pr -137-> Ce	2.8	2	1.5^h	β_0^+	1.8	2*			56D5	(54H3) report no 1.4^h Pr^{137}
59 78 — 58 79			see comment	no γ, β^+/ϵ	~ 0.2				56D5	activity. Theoretical β^+/ϵ ~ 0.5.
$5/2^+$ 5.8 $3/2^+$										

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ERROR	HALF-LIFE	TYPE	MEV	ERROR	%	METHOD	REF	COMMENTS
Xe --138--> Cs 54 84 55 83 0^+ > 6.2 3^-	2.4 or 3.7	3 3	17^m	β^-	2.4				55N4	
				$E\gamma/\beta = 1.3$					55N4	
Cs --138--> Ba 55 83 56 82 3^- 7.7 2^+ 0^+	4.83	5	32^m	γ	1.426 1.44	5 2		scin scin	56B6 55T2	complex γ spectrum. Establish consistent decay scheme by $\gamma\gamma$ (56B6) with $3.4\beta \sim 20$ per cent.
La --138--> Ba 57 81 56 82 $\underline{5^-}$ 16.9 2^+ 0^+	1.59 see comment	3	$\sim 10^{11}y$	γ	1.44 1.43	2* 1		scin scin	57G1 56T4	Calculate ϵ_L/ϵ_K =1.4(from K x ray/γ and assuming flourescent yield = 0.85) implies E_ϵ = 0.16 3 (56T4). Assumed ΔJ = 3,<u>no</u>.
				see Cs --138-->Ba						
				(x) (1.4γ)				scin	57G1 56T4	
				K x ray / gm / sec.						
					0.145 0.23	10			56T4 54S4	
La --138--> Ce 57 81 58 80 5^- 19.0 2^+ 0^+	1.02	2	$10^{11}y$	β^-	0.205	20*	100	scin	57G1	
				γ	0.81 0.81	1* 1*		scin scin	57G1 56T4	
Pr --138--> Ce 59 79 58 80 3, 5^+ ≥ 5.4 4^+ 2^+ 0^+	3.2 or 4.3	1 1	2.0^h	β^+	1.4	1		a, scin	54H3	0.8γ corresponds to 2^+ state in Ce^{138} (see above). Intensities of γ's suggest 1.05γ and 0.8γ may be in cascade.
				γ	0.30 0.80 1.05 1.4 1.7	1* 1* 2* 1 weak* 1 weak*		scin scin scin scin scin	54H3 54H3 54H3 54H3 54H3	
Cs --139--> Ba 55 82 56 83 7/2, $5/2^+$ ≥ 6.3 $7/2^-$	3.17	15	10^m	β_0^-	3.17	15*		a	54F5	
Ba --139--> La 56 83 57 82 $7/2^-$ 6.8 $5/2^+$ 7.4 $\underline{7/2^+}$	2.39	2	85^m	β_0^-	2.38	3*	15	s	54M13	other β, γ present (54M13).
				β_1^-	2.23	2*	66	s	54M13	
				γ	0.163 0.165	2* 5*		s scin	54M13 54N7	
				$(2.23\beta)(0.163\gamma)$					54M13	
Ce --139--> La 58 81 57 82 $3/2^+$ 5.8 $5/2^+$ $\underline{7/2^+}$	0.27	1	140^d	γ	0.1665	10*		s	54P6	
				ϵ_L/ϵ_K = 1.37		2		$\gamma\gamma$, xγ, xx	56K9	
				$E_\epsilon < 0.147(\epsilon_L/\epsilon_K)$					54P6	
				$E_\epsilon \sim 0.104$		6		ϵ_L/ϵ_K	56K9	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Pr --132--> Ce	2.0	1	4.5^h	β^+_o	1.0	1		a, scin	54H3	weak γ present (54H3)
59 80 58 81										
5/2$^+$ 5.5 3/2$^+$										
La --140--> Ce	3.78	3	40.2^h	β^-	2.20	2	8	s	56A2	very complex decay, but $\gamma\gamma$ studies
57 83 58 82				γ	1.601	10		s	56A2	permit consistent decay scheme.
3, 4$^-$ 9.1 2$^+$					1.60	20*		scin	55B15	(56K10, 55C8, 55B15, 55B16).
0$^+$					1.597	10*		s	55C8	
Ce --141--> Pr	0.577	2	33^d	β^-_o	0.574	3	25	s	55J4	β_o, β_1, F-K plot linear (55J4).
58 83 59 82				β^-_1	0.432	2	75	$\beta\gamma$s	55J4	
7/2$^-$ 7.1 7/2$^+$				γ_1	0.1416	3		s	56M6	
7.8 5/2$^+$					0.1449	7		s	55J4	
				$(\beta)(\gamma_1)$					55J4	
Pr --142--> Nd	2.160	15	19.2^h	β^-_o	2.12	10	97.2	a	55S9	β_o has ΔJ = 2,yes shape (54P3).
59 83 60 82					2.166	6	93	s	54P3	Energy match by (0.6β)(1.6γ), (55S9,
2$^-$ 7.1 2$^+$				γ	1.61	2		scin	55S9	54P3).
9.2^1 0$^+$					1.572	5*		s	54P3	
				$\beta\gamma$					55S9, 54P3	
Ce --143--> Pr	1.544	5	33.4^h	β^-	1.40	2	37	s	56M6	complex decay, $\beta\gamma$ + $\gamma\gamma$ studies
58 85 59 84				γ	0.0574	2		s	56M6	establish decay scheme with good
9/2$^-$ 7.8 7/2$^+$				(1.4β)(0.06γ), $\beta\gamma$, $\gamma\gamma$				scin	56M6	energy matches. (56M6).
5/2$^+$										
Pr --143--> Nd	0.929	5	14^d	β^-_o	0.93	1		s	56M6	
59 84 60 83				no γ		< 0.1 per cent				
5/2$^+$ 7.6 7/2$^-$								scin	55S6	
Sm --143--> Pm	3.5	2	9^m	β^+_o	2.6	3*		a	56S15	
62 81 61 82					2.3	3		a	56M7	
3/2$^+$ 5.2 5/2$^+$				noγ	(E_γ > 0.15)			scin	56M7	
Ce --144--> Pr	0.308	4	290^d	β^-_o	0.320	3*	60	s	57H2	Good energy match by ~ 0.17β +~0.14γ
58 86 59 85					0.309	5	76	s	56P7	(57H2, 56P7, 54C4).
0$^+$ 6.5 0$^-$					0.327	7	75	s	54C4	
Pr --144--> Nd	3.03	4	17.5^m	β^-_o	3.15	6*	99	s	57H2	some indication of non linear F-K
59 85 60 84					2.99	6*	99	s	56L8	plot (due to 0$^-$ ---> 0$^+$ nature of
0$^-$ 6.5 0$^+$					3.12	6*	99	s	54C4	transition with possible T-P inter-
										ference?)(56L8)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Sm $\xrightarrow{145}$ Pm; 62 83 → 61 84	> 0.1		~ 1^y	Pm K x ray					57C1	
$7/2^-$, $9/2^-$ → $5/2^+$, $7/2^+$				γ	0.0613			(K x ray)γ	57C1	
$5/2^+$, $7/2^+$				no $\gamma^\pm$, α					57C1	
Nd $\xrightarrow{147}$ Pm; 60 87 → 61 86	0.911	7	11.6^d	β^-	0.818	7		s	56E1	
$\underline{9/2^-}$ 7.6 $7/2^+$				γ	0.09071	1		crys	56K11	
					0.092	5		scin	55H7	
$5/2^+$				(0.818β)(0.091γ)				βγs	56E1	
Pm $\xrightarrow{149}$ Sm; 61 88 → 62 87	1.34?	1	50^h	β^-	0.97	5^*		s	54F6	With shell model assignment of $5/2^+$ to Pm^{149} it is difficult to explain the absence of a g. s. β^- transition.
$5/2^+$ 7.0 ?				γ	0.285	5^*		scin	54F6	
$\underline{7/2^-}$										
Pm $\xrightarrow{150}$ Sm; 61 89 → 62 88	5.3	1	2.7^h	β^-	3.00	5	20	s	54F6'	Complex γ. Level scheme constructed in accordance with γγ(54F6). Level scheme for Sm^{150} from $Sm^{149}(n, \gamma)$ in excellent agreement and has state at 2.30(55A7).
5, 6^- 8.2 ?				γγ				scin	54F6	
0^+				$3.0\beta^-$ decay is to level at 2.3		1		γγ	54F6	
Pm $\xrightarrow{151}$ Sm; 61 90 → 62 89	1.44	10	28^h	γ	0.335	10^*		scin	55H7	Complex γ. Constructs decay scheme by γγ with (1.1β)(0.3γ) for decay energy. (55H7)
$5/2^+$?> 6.9 $7/2^-$?				γγ					55H7	
$9/2^-$?										
Eu $\xrightarrow{152}$ Sm; 63 89 → 62 90	> 1.58	2	12.7^y	ε						Complex γ spectrum. Construct decay scheme by γγ and intensity arguments. Highest established level decays by $\gamma_1 + \gamma_2 + \gamma_3$ (56G8).
4, 5^+ ?				γ_1	0.122	1^*	59	s	56G8	
					0.12231	4	9	crys	56A5	
2^+				γ_2	0.244	2^*		s	56G8	
0^+					0.24464	8		crys	56A5	
				γ_3	1.210	20^*		scin	56G8	
				γγ					56G8,54S5	
Eu $\xrightarrow{152}$ Gd; 63 89 → 64 88	1.80	2	12.7^y	β^-	1.460	14^*	13	s	56G8	Complex decay. Construct consistent decay scheme by βγ and γγ arguments. Energy match provided by 1.0β + 0.41γ (56G8); 0.7β(54S5) + 0.78γ. No report of non linear F-K plot for 1.46β.
4, 5^+				γ	0.344	3^*		s	56G8	
					0.34434	23		crys	56A5	
11.9 2^+				(1.46β)(0.34γ)				scin	56G8	
0^+				γγ, βγ					56G8, 54S5	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Sm -153-> Eu 62 91 63 90	0.803	5	47^{h}	β_0^-	0.825 0.820	25* 10	27 40	scin s	56D6 55M15	Levels seen in coulomb excitation not in agreement with those seen in β decay. Complex decay. Consistent decay scheme with good energy matches possible.
$7/2^-$ 6.8 $7/2^+$				β_1^-	0.720	15	40	sβγ	55M15	
6.9 $5/2^+$				γ_1	0.10327 0.100 0.103	2 10* 2*		crys scin s	56A5 56D6 55M15	
				(0.72β)(0.10γ)					55M15	
Gd -153-> Eu 64 89 63 90	0.21	2	236^{d}	ε						
$7/2^-$ 7.2 $7/2^+$ $5/2^+$				γ	0.103 0.100 0.104 0.1031	1* 4* 5* 5		s scin scin s	56M8 56B7 56B8 54C5	
				from ϵ_L/ϵ_K estimate Q						
				Q	~0.22 (<0.23) 0.20 0.187 0.225				56M8 56B7 56B8 56B9	
Eu -154-> Sm 63 91 62 92 $8^{\pm}$? 2^+ 0^+	see comment		16^{y}							1.17γ assigned here because it is associated with Eu^{154} and does not fit the decay scheme of the Eu^{154} β^- branch (54S5).
Eu -154-> Gd 63 91 64 90	2.99	5	16^{y}	β^-	1.45 1.575	5 150		s s	54S5 48S1	Spin 6 for 1.54 state of Gd^{154} on basis of (1.42γ)(0.123γ)(θ) (56H7) and lifetime of state (54S5).
8^{-}_{+} ? ≥10 6^+				γ_1	1.415	30*		scin	54S5	
2^+				γ_2	0.12354 0.1234	9 5		crys s	56A5 54C5	
0^+				(1.4γ)(0.12γ)					56H7, 54S5	
Eu -155-> Gd 63 92 64 91 $5/2^+$ 8.2 see comment	0.247	3	1.7^{y}	β_0^-	0.240	10	21	scin	56D6	0.15β + 0.102γ provide fair energy match. No(0.24β)γ (56D6). g. s. spin has been reported as $3/2^+$(56S16) and ≥ $5/2^+$ (54M14). The decay data makes no choice between $3/2^+$, $5/2^+$, $7/2^+$.
Eu -156-> Gd 63 93 64 92 9.1 2^+ 0^+	2.54	5	15.4^{d}	β^-	2.45	5		s	56B10	decay scheme shows 2.45β to 0.09 state in Gd^{156}. (56B10)
				γ	0.08897	1		crys	56B10	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
$Tb \xrightarrow{156} Gd$ 65 91 → 64 92; 0^+	see comment		5^h	no $\gamma^{\pm}$				scin	55H8	5^h Tb^{156} is assigned as g. s. see comments in β^- decay. Severe disagreement with previous data.
				no Gd X ray				pc	55H8	
				5.2^d Tb Gd X ray, no $\gamma^{\pm}$				pc	55H8	
$Tb \xrightarrow{156} Dy$ 65 91 → 66 90; > 5.4 0^+	see comment		5^h	β^-	0.14	1*		a	55H8	5.2^d Tb decays by ϵ, $0.6\beta^-$ and hard γ. Hence 5^h is g. s. β^-/ϵ very small (55H8).
				e^- and Dy L x ray				pc	55H8	
				no hard γ				a	55H8	
$Gd \xrightarrow{159} Tb$ 64 95 → 65 94; $3/2^-$ >6.4 $3/2^+$	0.948	10	18^h	β^-_0	0.948	10	76	s	55M16	$(0.6\beta)(0.36\gamma)$ provides good energy match to β_0 (55M16). No $(0.36\gamma)(0.05\gamma)$ thus indicates β_0 is the g. s. transition.
				$\beta\gamma$, $\gamma\gamma$					55B17, 55M16	
$Tb \xrightarrow{160} Dy$ 65 95 → 66 94; ? 8.7 ?; 11.9 2^+; 0^+	1.82	1	73^d	β^-_1	~1.700				57C1	Very complex decay. (56B11, 54K5) have constructed decay schemes incorporating most of the well known data.
					1.711	7	0.3	s	56B11	
				β^-_2	0.861	2	35	s	56B11	
					0.851	10	30	s	54K5	
				γ_1	0.0880	5		s	56B11	
				$\beta\gamma$, $\gamma\gamma$					56B11	
$Tb \xrightarrow{161} Dy$ 65 96 → 66 95; ? 7.1 ?; $5/2^-$	0.562 or 0.596 see comment	10 / 4	7.15^d	β^-	0.571	4	14	s	56S17	Complex β spectrum makes analysis open to wide error. Several decay schemes proposed. A level scheme with excellent energy matches between γ rays (56C3) is fit very closely by the β groups from the most careful β spectrum analysis (56S17), if it is assumed that $0.57\beta^-$ goes to 0.026 level in Dy^{161}. Under these conditions g. s. spin of Dy^{161} is 7/2(53M2).
					0.531	10	6	s	56C3	
					0.550	10		s	55B17	
				$\gamma\gamma$	56S17, 56C3, 56B12, 56B13, 55B17.					
				$\beta\gamma$					56B12, 55B17	
$Er \xrightarrow{161} Ho$ 68 93 → 67 94; 7/2, $9/2^-$ ≥ 5.8 ?; $7/2^+$	≥ 2.1 see comment	1	3.1^h	β^+ see comment	1.2	1		a	55N2	However, (54H4) reports no $\gamma^{\pm}$.
$Ho \xrightarrow{164} Dy$ 67 97 → 66 98; 2^+; 0^+	0.11 ≤ Q ≤ 1.4		36.7^m	γ_1	0.0728	5		s	54B8	However, 0.046γ may be possible isomeric transition t long lived Ho^{164} g. s. (54B8, 52B1).
				γ_2	0.0373	5		s	54B8	
				$\epsilon_K/\beta^+ > 2 \times 10^3$				s	54B8	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Ho -164-> Er	0.98	3	36.7^m	β_o^-	0.99	3	25	s	54B8	$\beta\gamma$ and β spectrum suggest strong
67 97 68 96					0.9	1*		a	54H5	
1^- 2^+				$\beta\gamma$					54B8	β group in coincidence with 0.09γ (54B8). See comment for Ho-164->Dy.
5.9 0^+										
Ho -166-> Er	1.853	4	27.3^h	β_o^-	1.854	5	52	s	55G4	Scin and $\beta\gamma$ studies show that Q for
67 99 68 98				β_1^-	1.771	7	48	$s\beta\gamma$	55G4	30y Ho^{166} decay > 2.1 (55M17). F-K
0^- 9.4^1 2^+				γ_1	0.0803	2		s	55G4	plot for β_o is linear, for β_1 has
8.4 0^+				$\beta_1\gamma_1$					55G4	ΔJ = 2; yes shape (55G4).
Ho -167-> Er	1.0	1	3^h	β_o^-	1.0	1	50	scin	55H9	
67 100 68 99				γ/β ~ 0.25					55H9	
$7/2^+$ 6.4 $7/2^-$										
Tm -167-> Er	> 0.72	1	9.6^d	γ	0.720	10*		scin	55N2	lower energy γ's present (55N2).
69 98 68 99										
$1/2^+$?										
$7/2^-$										
Yb -167-> Tm	3.4	2	74^m	β_o^+	2.4	2		s	55N2	
70 97 69 98										
? ?	or		or							
or	>0.118	2	18.5^m							
$1/2^-$ > 5.6 $1/2^+$										
Er -169-> Tm	0.339	2	9^d	β^-	0.330	10	85	s	56H8	85 per cent β^- to g. s., 15 per cent
68 101 69 100					0.340	2		s	56B14	to 0.008 state. (56H8).
$1/2^-$ 6.9 $3/2^+$				γ	0.00842	5		s	56H8	
6.1 $1/2^+$										
Yb -169-> Tm	> 0.473	1	30.6^d	ϵ						
70 99 69 100				many γ	Highest state reached					
? $3/2^+$					0.47291	10*		crys	56H9	
$1/2^+$					0.474*	1*		s	56M9	
					0.4732*	7*	$\gamma\gamma$	s	56C4	
					0.473	5*	$\gamma\gamma$	scin	56K12	
					0.474	5*	$x\gamma$, $\gamma\gamma$	scin	55J5	
Tm -170-> Er	> 0.05		127^d	Er K x ray					56D7	ϵ/β^- ~ 1/100 (56D7)
69 101 68 102										
1^- 0^+										

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
$Tm \xrightarrow{170} Yb$ 69 101 → 70 100	0.966	3	127^d	β_o^-	0.950	5	76	scin	56B15	F-K plot for both β_o and β_1 is linear (54P3). $\beta_1\ \gamma_1$ observed (56B15, 54P3).
					0.970	2	78	s	54P3	
1^- 2^+				β_1^-	0.867	5	24	scin$\beta\gamma$	56B15	
					0.886	9	22	s$\beta\gamma$	54P3	
9.0 0^+				γ_1	0.084229	41		crys	56D7	
					0.08426	2		crys	55H10	
$Er \xrightarrow{171} Tm$ 68 103 → 69 102	1.48	1	7.8^h	β_1^-	1.52	5*	6	scin$\beta\gamma$	57J4	$\gamma\gamma$ and $\beta\gamma$ studies fit γ's reported by (51K1) into decay scheme. Level spacings in agreement with a K = 1/2 rotational band (57J4, 56C5, 56K11).
					1.45	2*	3	s	56C5	
$5/2^-$ 6.4 $7/2^-$				β_2^-	1.10	3*	94	scin$\beta\gamma$	57J4	
8.1 $3/2^+$					1.05	1*	97	s	56C5	
$1/2^+$				$\beta\gamma$, $\gamma\gamma$			57J4, 56C5, 56K11			
$Tm \xrightarrow{171} Yb$ 69 102 → 70 101	0.103	2	680^d	β_o^-	0.103	2		s	55B18	
$1/2^+$ 6.4 $1/2^-$				no γ					55B18	
$Tm \xrightarrow{172} Yb$ 69 103 → 70 102	> 1.79	5	63.6^h	β^-	1.5	1*		a	56N4	other, lower energy γ's noted (56N4)
$?^-$ >8 0^+				γ	1.79	5*		scin	56N4	
$Yb \xrightarrow{175} Lu$ 70 105 → 71 104	0.467	3	4.2^d	β_o^-	0.471	3	80	s	56C4	Complex decay. Good energy matches by (0.07β)(0.4γ) and (0.35β)(0.11γ). Consistent decay scheme by (56H9, 56C4, 55M18, 55D5). Level spacings in Lu^{175} in agreement with a K = 7/2 rotational band.
$7/2^-$ 6.4 $7/2^+$					0.468	5	80	s	55M18	
					0.463	3	87	s	55D5	
$Hf \xrightarrow{175} Lu$ 72 103 → 71 104	> 0.541		70^d	ϵ						Complex decay. Level spacings in Lu^{175} in agreement with K = 7/2, 5/2 bands. β decay energy systematics (54W2) suggest Q ~ 0.9.
$5/2^-$ $7/2^+$				many γ. Highest state reached.						
					0.5407	5*		s, crys	56H9	
				$\gamma\gamma$					56H9, 55M18	
$Lu \xrightarrow{176} Yb$ 71 105 → 70 106	> 0		3 x 10^{10}y	ϵ				pc	54D4	ϵ/β^- = 0.03 (54D4)
9, 10^- 0^+										
$Lu \xrightarrow{176} Hf$ 71 105 → 72 104	1.00	2	3 x 10^{10}y	β^-	0.43	2*		pc	54D4	
9, 10^- 18.4 6^+				γ_1	0.19	1*		scin	54D4	
4^+					0.203	4		scin	54A3	
2^+				γ_2	0.31	1*		scin	54D4	
0^+					0.306	6*		scin	54A3	
				γ_3	0.089	2*		scin	54A3	
				$\gamma\gamma$					54D4	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Yb -177-> Lu 70 107 71 106 $5/2^-$ 6.2 $5/2^+$	1.38	3	1.8^h	β_0^-	1.30	5	88	s	55D5	Decay scheme established with good energy matches (e. g. (0.16β⁻) (>0.7γ), (56M10)).
				β_1^-	0.16	2		scinβγ	56M10	
				γ	1.24 1.228	2 5		scin scin	56M10 56C4	
				βγ, γγ					56M10, 56C4, 55D5	
Lu -177-> Hf 71 106 72 105 $5/2^+$ 6.8 $7/2^-$	0.494	5	6.8^d	β_0^-	0.497	2	90	s	55M19	Decay scheme establishes β_0 as g. s. transition (55M19).
Ta -177-> Hf 73 104 72 105 $5/2^+$ > 6.5 $7/2^-$	1.15	5	2.2^d	ε					57M1	Several γ present (57M1). ϵ_L/ϵ_K suggest Q = 1.15 5 (57M1)
Re -177-> W 75 102 74 103	> 1.4		17^m	β^+	> 0.4			s	57H3	
Ta -178-> Hf 73 105 72 106 0^+	see comment			no γ_-^+				scin	56B16	15^d β^+, 1.5γ previously reported here.
				no 1.5γ				scin	56B16	
W -178-> Ta 74 104 73 105 0^+	> 0.07		22^d	ε, K x ray				scin	56B16	
Re -178-> W 75 103 74 104 ≥ 5.2 0^+	≥ 4.1	1	15^m	β^+	3.1	1^*		s, a	57H3	
W -179-> Ta 74 105 73 106 $7/2^-$? $5/2^+$?	> 0.03	1	40^m	γ	0.030		9^*		56R1	
Ta -180-> Hf 73 107 72 108 1^- 8.1^h 6.8 ? 0^+	see comment > 0.093		$>10^{12}y$ or 8.1^h							Ta^{180}/Ta^{181} = 12.3/99,988 in natural Ta (55W5). t estimated from abundance of W^{180} in Ta ores (55E1) 8.1^h previously reported here may not be g. s.
Ta -180-> W 73 107 74 106 1^- 8.1^h 0^+	see comment ≤ 0.70	2	$> 10^{12}y$ or 8.1^h	see comment in Ta -180-> Hf						

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Re -180-> W 75 105 74 106	≥ 2.1	1	2.4^m	β^+	1.1	1	20	a	55F7	0.106γ also present (55F7)
? > 3.5	or		or	γ	0.880	40		scin	55F7	
or	≥ 2.9	2	20^h	β^+	1.9	2		s	57H3	
> 6.4 0^+										
W -181-> Ta 74 107 73 108	0.16	1 <		γ_1	0.152	10*		scin	56D8	γ intensity ~ 10^{-3}/K x ray (56D8).
7/2⁻ $\underline{7/2^+}$	Q < 0.22	2		γ_2	0.136	10*		scin	56D8	(56B1) reports no γ and estimates Q
				(L x ray) γ; no (K x ray) γ					56D8	from K x ray/L x ray. 0.136γ appears, 0.152γ does not appear in Hf-181->Ta. (56B17).
Ta -182-> W 73 109 74 108	1.732	10	111^d	β^-	0.51	1		s	55M20	Extremely complex γ spectrum (~ 30
> 8 2^+ 0^+				γ	1.2220 1.222 1.222	2* 1 1		s s s	56B18 55M20 54F7	γ's). Decay schemes consistent with 0.51β to 1.22 level (56B18, 55M20, 54F7).
Hf -183-> Ta 72 111 73 110	2.2	2	64^m	β^-	~1.4	2*		a	56G11	
1/2⁻ ≥6.0 1/2⁺ 5/2, 7/2⁺				Eγ/β-	0.8			GM	56G11	
Ta -183-> W 73 110 74 109	1.068	10	5^d	β^-	0.615 0.6	10 1*	>95	s scin	55M20 55P4	Extremely complex γ spectrum (55M20,
5/2, 7/2⁺				0.615β⁻ to state at						55P4). Decay scheme established by
6.9 7/2⁻ $\underline{1/2^-}$				0.45308		10*		crys	55M20	combination principle. Weak, higher energy β⁻ present with E_β < ~1 (55M20).
Ta -184-> W 73 111 74 110	≥ 1.26	7	8.7^h	β^-	1.26	7	70	scin	55B19	9 γ's reported (Eγ 0.11 to 1.18).
?⁻ 6.9 0^+										(55B19)
Ta -185-> W 73 112 74 111	≥ 1.72	7	49^m	β^-	1.72	7		scin	55P4	4 γ's reported (Eγ 0.06 to 0.235)
7/2⁺ ≥6.3 7/2⁻										(55P4).
W -185-> Re 74 111 75 110	0.428	3	73^d	β_o^-	0.435 0.426	5 3	 90	s s	57D1 55B20	Weak γ reported (56B19, 55B20, 55K10).
7/2⁻ 7.5 $\underline{5/2^+}$										Possible 0.37β⁻ (55F20) + 0.06γ provides energy match.
										1

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Os —185→ Re 76 109 — 75 110 $1/2^-$ — $5/2^+$	> 0.879	1	97^d	γ	0.8789 0.875	9* 10*		s scin	56N5 55P5	Other γ's reported (56N5, 55P5). Highest state reached in Re^{185} at ~ 0.88 (55P5).
				γγ_					55P5	
Ta —186→ W 73 113 — 74 112 $2,,3^-$ ≥ 6.2 2^+ 0^+	>2.3	2	10^m	β^-	2.2	2		scin	55P6	8 other γ's observed. 0.125γ is from 0.12 state in W^{186} (see Re —186→ W).
				γ	0.125			scin	55P6	
				(2.2β)γ				GM	55P6	
Re —186→ W 75 111 — 74 112 1^- — 2^+ 0^+	>0.123	1	3.7^d	ε						ϵ/β^- ~ 3/100 (56J8)
				γ	0.12267 0.1229	21 4		s s	56P8 56J8	
				no β^+					56P8	
Re —186→ Os 75 111 — 76 110 1^- 8.0 2^+ 7.7 0^+	1.067	2	3.7^d	β^-_0	1.0715 1.064 1.060 1.063	13 2 5* 5	76 76 73 76	s s s s	56P8 56J8 56A2 54K6	(0.93β)(0.137γ) provides energy match (56P8, 56J8, 54K6). (56P8) reports F-K plots for β_o and for β_1 are non linear. (54K6) reports both linear.
				β^-_1	0.9335 0.926	30 5	24 24	sβγ sβγ	56P8 54K6	
				γ	0.13719 0.1372 0.13722	7 2 3		s s crys	56P8 56J8 55R9	
W —187→ Re 74 113 — 75 112 $1/2^-$ 9.1[1] $5/2^+$	1.312	6	24^h	β^-_0	1.320	20		s	57D1	0.62β + 0.69γ provide good energy match. F-K plot for β_o has ΔJ = 2,yes shape (57D1)
Re —187→ Os 75 112 — 76 111 $5/2^+$ — $1/2^-$	~ 0 see comment		<10^{11}y or >10^{16}y	β^-	<0.008			a	54S6	(54D5) reports no β with E_β > 0.001. Measurement of Os^{187} spin (54M15) and shell model assignment of parity change leads to ΔJ = 2;yes transition and suggests Q ~ 0.
Ir —187→ Os 77 110 — 76 111 $3/2^+$ — $1/2^-$	≥ 0.435	10	14^h	γ±(?)				scin	55S10	other γ's present $\gamma^\pm$ may be due to Ir^{188} contaminant (55S10).
				γ	0.435	15		scin	55S10	
Re —188→ Os 75 113 — 76 112 1^- 8.6 2^+ 8.1 0^+	2.114	6	17^h	β^-_0	2.15 2.116 2.01	4 2 2*	79 79	s s s	56P9 56J8 56A2	(2.0β)(0.155γ) provides energy match (56P9).
				βγ					56P9	
				γγ					56P9, 56J8	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Ir $\xrightarrow{188}$ Os 77 111 76 112 0^+	> 0.625	10	41^h	no $\gamma^{\pm}$				scin	55S10	0.15γ (55S10, 55N5) and 0.475γ (55S10) also present.
				γ	0.625	15		scin	55S10	
Pt $\xrightarrow{188}$ Ir 78 110 77 111 0^+ ?	> 0.4	1	10^d	no γ± (?)				scin	55S10	other γ's (55S10), e^- (55N5) present.
				γ ~ 0.4		1^*		scin	55S10	
Ir $\xrightarrow{189}$ Os 77 112 76 113 $3/2^+$ $\underline{3/2^-}$	≥ 0.245	10	11^d	no $\gamma^{\pm}$				scin	55S10	
				γ	0.245	10		scin	55S10	
Pt $\xrightarrow{189}$ Ir 78 111 77 112 $1/2^-$ $3/2^+$	> 0.14	1	10.5^h	no $\gamma^{\pm}$ (?)				scin	55S10	possible γ(0.55 to 0.70)(55S10).
				γ	0.14	1		scin	55S10	
Au $\xrightarrow{189}$ Pt 79 110 78 111 $3/2^+$ $1/2^-$	> 0.29	1	42^m	no $\gamma^{\pm}$ (?)				scin	55S10	0.135γ, possible > 0.8γ present (55S10).
				γ	0.29	1		scin	55S10	
Re $\xrightarrow{190}$ Os 75 115 76 114 5^-(?) ≥5.5 4^+ 2^+ 0^+	1.7 Q ≤ 3.2	3≤ 3	2.8^m	β^-	1.7	3		a	55A8	Several γ (0.19 to 0.83) present (55A8).
				Eγ/β ~ 1.5					55A8	
Ir $\xrightarrow{190}$ Os 77 113 76 114 9, 10, 11^+ 3^h 3, 4, 5^- 11^d 10^m 10^- 2^+ 0^+	1.33 Q < ~2	3<	11^d	From 3^h Ir^{190}						3^h Ir decay leads to level scheme and spin assignments in Os^{190}. Absence of β^+ from 11^d Ir leads to estimate of upper bound to Q(55A9).
				β^+	2.04	7		scin	55A9	
				From 11^d Ir^{190}						
				no β^+				scin	55A9	
				γ	1.33	3^*		scin	55A9	
Pt $\xrightarrow{191}$ Ir 78 113 77 114 $1/2^-$ $\underline{3/2^+}$	≥ 0.625	2	3^d	ε						Complex γ (55T7, 55S10, 54C6). γγ establishes highest level reached in Ir^{191} at ~ 0.62 (56P10).
				γ	0.626 0.623	3^* 3^*		s s	55T7 54C6	
				γγ					56P10	
Au $\xrightarrow{191}$ Pt 79 112 78 113 $3/2^+$ $1/2^-$	> 0.60	2	3^h	ε						other γ's (0.14 to 0.48) also present. (55S10).
				γ	0.60	2		scin	55S10	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Ir -192-> Os 77 115 76 116	1.065	1<	74^d	ϵ			4.3 3.5		56B20 55B21	Many γ's. Level scheme with good energy matches indicate that highest level reached in Os^{192} decays by $\gamma_1 + \gamma_2 + \gamma_3$ as well as by direct g. s. γ (56A2, 55B21). Upper bound to Q is estimated from lack of β^+ to 4^+ 0.69 level (55B21).
5^- 4^+	Q < ~ 2			no β^+				crys	55B21	
2^+				γ_1	0.3747 0.374	5 2		s,crys s	55B21 54J4	
0^+				γ_2	0.48475 0.4840 0.4844	23 5* 2		crys s s	55B21 55G5 54J4	
				γ_3	0.20575 0.2054	4 2		crys s	55B21 54J4	
Ir -192-> Pt 77 115 78 114	1.452	6	74^d	β^-	0.673 0.672	10 10	48	s s	55B21 54J4	** γ_1 and γ_2 energies are each weighted average of 4 spectrometer and crystal measurements (55B21, 55R9, 55G5, 54J4). $\gamma\gamma$ and Σ scin establish decay scheme. Q from $0.67\beta + 0.31\gamma + 0.47\gamma$.
5^- 8.3 4^+				γ_1	0.3165**	1				
2^+				γ_2	0.4679**	1				
0^+				$\gamma\gamma$					56K10, 55B2, 55T8	
				Σ scin					54L6, 54P7	
Os -193-> Ir 76 117 77 116	1.134	9	31.5^h	β_0^-	1.136	5	<100	s	56N5	Complex γ and β. Good energy matches in decay scheme involving 5 excited states of Ir^{193} (56N5).
$1/2^-$ ≥ 7.3 $3/2^+$										
Pt -193-> Ir 78 115 77 116	0.013 <		≤ 500^y	ϵ_L/ϵ_K					56N6	$\epsilon_L/\epsilon_K > 1000$ (56N6).
$1/2^-$ $3/2^+$	Q < 0.080			no β^+, β^-, γ, K x ray				scin	56N6	
Hg -193-> Au 80 113 79 114	> 1.646	4	11^h	β^+				s	56B24	Very complex γ. Other γ with E_γ < 1.6 present (56B24).
$1/2^-$ $3/2^+$				γ	1.646	4		s	56B24	
Ir -194-> Pt 77 117 78 116	2.233	13	19^h	β_0^-	2.236	10	66	s	54J4	Complex γ and β. Decay scheme with good energy matches (55M21, 54J4).
1^- 8.2 0^+				$\beta\gamma$, $\gamma\gamma$					55M21	
Au -194-> Pt 79 115 78 116	2.566	14	39^h	β_0^+	1.55	2		s	56T5	$(1.21\beta)(0.33\gamma)$, complex γ. (56T5).
1^- ≥7.7 0^+				β_1^+	1.21	2		s$\beta\gamma$	56T5	
				γ	0.3267	5		s	56T5	
				$\beta\gamma$, $\gamma\gamma$					56T5	

TABLE I (*continued*).

Disintegration	Adopted Q MEV	Adopted Q Error	Half-life	Decay data Type	Decay data MEV	Decay data Error	Decay data %	Decay data Method	Decay data Ref.	Comments
$\mathrm{Au}_{79\ 116} \xrightarrow{195} \mathrm{Pt}_{78\ 117}$	0.27	2	185^d	ϵ					54B9	Several γ's present. (56P11, 55B23. 54B9). ϵ_L/ϵ_K implies E_ϵ = 0.14 2 (54B9).
$3/2^+$ ≥6.8 $1/2^-$				ϵ_L/ϵ_K to 0.13 level	0.58	14			54B9	
$\mathrm{Hg}_{80\ 115} \xrightarrow{195} \mathrm{Au}_{79\ 116}$	≥1.21	1	9.5^h	γ_1	1.150	10*			55J6	Other γ's present. Coincidence studies support level scheme with highest level at 1.21 . (55J6)
$3/2^-$ $3/2^+$				γ_2	0.0614	5*			55J6	
				$(0.06e^-)(1.15\gamma)$					55J6	
$\mathrm{Tl}_{81\ 114} \xrightarrow{195} \mathrm{Hg}_{80\ 115}$	≥ 0.037	1	1.2^h	ϵ						higher energy γ may be present. (57A1).
$1/2^+$ $3/2^-$				γ	0.0370	2*		s	57A1	
$\mathrm{Pb}_{82\ 113} \xrightarrow{195} \mathrm{Tl}_{81\ 114}$	see comment		see comment	From 17^m Pb^{195}						Suggest 17^m state is <u>not</u> Pb^{195} g. s. (57A1). If γ_1, γ_2, γ_3 in cascade, Q > 0.94 - I.T. γ(unobserved) (57A1).
17^m $13/2^+$				γ_1	0.3931	4*		s	57A1	
? $11/2^-$				γ_2	0.4475	5*		s	57A1	
$1/2^+$				γ_3	0.0990	2*		s	57A1	
$\mathrm{Au}_{79\ 117} \xrightarrow{196} \mathrm{Pt}_{78\ 118}$	> 0.685	1	5.6^d	γ_1	0.3310	4*		s	56T6	β decay energy systematics (54W2) suggest Q ~ 1.3.
2, 3^- 0^+					0.330	5*		scin	55B23	
				γ_2	0.3540	4*		s	56T6	
					0.360	5*		scin	55B23	
				$\gamma_1\gamma_2$					56T6,55B23	
$\mathrm{Au}_{79\ 117} \xrightarrow{196} \mathrm{Hg}_{80\ 116}$	0.70	2	5.6^d	γ	0.430	5*		scin	55B23	β^- branching 12 per cent (56T6).
2, 3^- 6.8 2^+				$\beta^-(0.43\gamma)$					55B23	
0^+										
$\mathrm{Tl}_{81\ 115} \xrightarrow{196} \mathrm{Hg}_{80\ 116}$	> 0.425	1	2.4^h	γ_1	0.4246	5*		s	57A1	
2,3 2^+					0.426	2		s	55A1	
0^+										
$\mathrm{Pb}_{82\ 114} \xrightarrow{196} \mathrm{Tl}_{81\ 115}$	> 0.420	1	37^m	γ_1	0.1668	2*		s	57A1	(e^-_I) (e^-_2) observed (57A1), γ's may be in Pb^{196} metastable state (57A1).
0^+ 2, 3				γ_2	0.2529	3*		s	57A1	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Pt -197-> Au	0.75	1	18^h	β_1^-	0.670	14*		scin$\beta\gamma$	56P10	no $\gamma\gamma_3$(56P10).
78 119 79 118				β_2^-	0.479	10*		scin$\beta\gamma$	56P10	
1/2⁻ 6.2 1/2⁻				β_3^-	0.468	10*		scin$\beta\gamma$	56P10	
3/2⁺				γ_1	0.077	2*		scin	56P10	
				γ_2	0.191	4*		scin	56P10	
				γ_3	0.279	6*		scin	56P10	
				$\beta_1\gamma_1$, $\beta_2\gamma_2$, $\beta_3\gamma_3$					56P10	
Hg -197-> Au	>0.268	1	65^h	ϵ						
80 117 79 118				γ_1	0.0776	5*		s	55J6	
1/2⁻ 1/2⁺				γ_2	0.1918	5*		s	55J6	
3/2⁺										
Tl -197-> Hg	> 0.611	2	2.8^h	ϵ						Strong 0.15γ (55K11, 55A10). Other
81 116 80 117				γ	0.611	2*	weak	s	55A10	γ's present (55A10).
1/2⁺ 1/2⁻										
Pb -197-> Tl	> 0.759	1	$<<42^m$	ϵ						e^- e^- establishes γ_1 γ_2 γ_3 in
82 115 81 116				γ_1	0.3867	4*		s	57A1	cascade and follow ϵ from 42^m
					0.388	1*		s	55A10	
13/2⁺ 42ᵐ										Pb^{197} (57A1). γ_4 is I. T. γ to
5/2⁻ 11/2⁻				γ_2	0.3848	4*		s	57A1	short lived Pb^{197} g. s. (57A1)
1/2⁺				γ_3	0.2219	3*		s	57A1	
				γ_4	0.2340	3*		s	57A1	
Au -198-> Pt			2.694^d	ϵ_K			$< 10^{-2}$		56B20	
79 119 78 120										
2⁺ 0⁺										
Au -198-> Hg	1.373	2	2.694^d	β^-	0.960	2		s	56P8	
79 119 80 118					0.967	5		s	54P3	
2⁺ 7.4 2⁺				γ	0.41176	19		elec s	56C6	
11.5 0⁺										
Tl -198-> Hg	>1.44	5	5.3^h	ϵ						other γ's present (55K11).
81 117 80 118				γ	1.44	5*		scin	55K11	
2⁺										
0⁺										

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
$_{82}Pb_{116} \xrightarrow{198} {}_{81}Tl_{117}$ 0^+	> 0.763	1	2.3^h	ε						Other γ's present (57A1, 55A10). $(e_1^-)(e_2^-)$(57A1). Previously reported 25^m isomer (54P8) not found (55A10).
				γ_1	0.3978	4*		s	57A1	
				γ_2	0.3655	4*		s	57A1	
$_{78}Pt_{121} \xrightarrow{199} {}_{79}Au_{120}$ $5/2^-$ >6.4 $\underline{3/2^+}$	1.78	6	30^m	β_0^-	1.7	2		a	56L9	aβγ and γγ suggest decay scheme with 1.1β to 0.63 level, 0.8β to 0.96 level (56L9).
				β_1^-	1.3	1		aβγ	56L9	
				γ_1	0.197	1*		s	56L9	
				γ_2	0.318	1*		s	56L9	
				βγ, γγ					56L9	
$_{79}Au_{120} \xrightarrow{199} {}_{80}Hg_{119}$ $\underline{3/2^+}$ 7.8 $\underline{1/2^-}$	0.455	3	3.15^d	β_0^-	0.460	10*	6	s	55H11	0.3β + 0.16γ provides good energy match.
				β_1^-	0.302	5*	69	s	55H11	
$_{82}Pb_{117} \xrightarrow{199} {}_{81}Tl_{118}$ $5/2^-$ $1/2^+$	> 0.721	2	90^m	ε						Weak 0.72γ suggests γ_1 and γ_2 are in cascade (55A10).
				γ_1	0.354	1*		s	55A10	
				γ_2	0.367	1*		s	55A10	
$_{78}Pt_{122} \xrightarrow{200} {}_{79}Au_{121}$ 0^+ 0, 1^-	> 0		11.5^h	p	Au^{200}			chem	57R1	Previously reported ~ 82^d activity is 76^d Ir^{192} (56W3).
$_{79}Au_{121} \xrightarrow{200} {}_{80}Hg_{120}$ 0, 1^- ≥ 6.9 0^+	2.3	1	48^m	β_0^-	2.2	1*		a	57R1	1.23 γ present (57R1).
$_{81}Tl_{119} \xrightarrow{200} {}_{80}Hg_{120}$ 0^+	> 2.137	2	26^h	ε						Complex γ(56G9, 55B24).
				complex γ. Highest level reached						
				2.1367		20*		e^-e^-,s	56G9	
$_{82}Pb_{118} \xrightarrow{200} {}_{81}Tl_{119}$ 0^+	> 0.416	1	18^h	ε						Complex γ(56G9, 55B24).
				Complex γ. Highest level reached						
				0.416		1*		e^-e^-, s	56G9	
$_{81}Tl_{120} \xrightarrow{201} {}_{80}Hg_{121}$ $1/2^+$ $\underline{3/2^-}$	> 0.167	1	3^d	ε						Complex γ. (55B24). e^-e^- studies suggest highest level reached is 0.167(55B24).
				γ	0.1672	10*		s	55B24	
				e^-e^-					55B24	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF	COMMENTS
Pb 82 119 -201-> Tl 81 120 1/2$^+$	> 1.53	1	9.4^h	ε γ e$^-$e$^-$	 1.099 	 5* 		 s 	 55B24 55B24	Complex γ. (55B24). e$^-$e$^-$ studies suggest highest level reached is 1.53(55B24).
Tl 81 121 -202-> Hg 80 122 0$^+$	0.54 Q < 0.75	3<	12^d	ε γ ϵ_L/ϵ_K 	 0.4406 = 0.90 = 0.43	 4 27 		 s 	 56K13 56K13 54H7	Limits on Q are from ϵ_L/ϵ_K and possible ft values (56K13).
Pb 82 120 -202-> Tl 81 121 0$^+$	0.015 Q < 0.085	<	~ 3x10^{5y}	no ϵ_K ϵ_L/ϵ_M	 ~ 1.5				54H7 54H7	Limits on Q from ϵ_K ~ 0 and existence of L capture.
Hg 80 123 -203-> Tl 81 122 5/2$^-$ 6.4 3/2$^+$ > 10.6 1/2$^+$	0.486	3	47^d	β$^-$ γ 	0.195 0.214 0.219 0.210 0.2783 0.279 0.279 0.279	2 2 10 2 5* 2 2* 2		s s s s s s s s	56W1 55M22 54W3 54T3 56W1 55M22 54W3 54T3	no 0.49β(56W1, 54W3). sβγ confirms decay scheme (55M22).
Tl 81 123 -204-> Hg 80 124 2$^-$ 0$^+$	0.34	1	4.26^y	ε continuous γ spectrum E_{max} ϵ_L/ϵ_K	 0.293 = 0.33	 20 		 scin xγ 	 56J4 54J2	K binding energy = 0.0833.
Tl 81 123 -204-> Pb 82 122 2$^-$ 8.9^1 0$^+$	0.765	2	4.26^y	β_o^- 	0.765 0.762 0.766	10 5 2		s s s	55Y2 54F2 54K7	β_o has ΔJ = 2;yes shape(55Y2, 54F2, 54K7).
Bi 83 121 -204-> Pb 82 122 0$^+$	>3.45	5	11.6^h	ε γ_1 γ_2 γγ	 0.981 1.20 	 5* 4* 		 s scin 	 56W4 56W4 56W4	Complex γ. (0.98γ)(1.2γ) precedes 1.266 0.26μs level in Pb^{204}(56W4).
Pb 82 123 -205-> Tl 81 124 5/2$^-$ 1/2$^+$	0.015 Q < 0.085	<	5 x 10^{7y}	ε L, L x ray no K x ray				scin scin	56H10 56H11,54C7	ϵ_L/ϵ_K > 200 from reported limits on τ_{ϵ_L} and τ_{ϵ_K} (56H10, 56H11).
Bi 83 122 -205-> Pb 82 123 9/2$^-$ ≥6.1 7/2$^-$ 5/2$^-$	2.65	3	14.5^d	β$^+$ γ γγ	0.925 0.7033 	30 8* 		s s 	56S19 56S19 56S19, 56G9	Extremely complex γ spectrum (56S19, 56G9). Consistent decay scheme constructed (56S19, 56G9). (γ_-^+)(0.70γ) demonstrated (56S19).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Bi $\xrightarrow{206}$ Pb 83 123 → 82 124 6^+ 145$^{\mu s}$ 7^- 0^+	3.404 Q < 3.7	3<	6.4^d	ε γγ, $e^-\gamma$, e^-e^- β^+ < 1/2600				scin, s s	54A4 54A4	Extremely complex γ spectrum. Coincidence studies and precise γ energy measurements fit decay scheme deduced from shell model (54A4). Highest level at 3.4035 (54A4)
Bi $\xrightarrow{207}$ Pb 83 124 → 82 125 $9/2^-$ $7/2^-$ $1/2^-$	2.357 Q ≤2.431	5< 5	8^y	ε γ_1 γ_2 Only L capture to 2.34 level	 0.5696** 1.771**	 2 5			55A1, 55L10	** γ_1 is weighted average of 5 spectrometer and 2 scin measurements (56B21, 55M23, 55Y3, 55A1, 55L10, 54P9, 54W4). ** γ_2 is weighted average of 1 spectrometer and 2 scin measurements (55A1, 55L10, 54P9).
Tl $\xrightarrow{208}$ Pb 81 127 → 82 126 5^+ 5.6 5^- 3^- 0^+	4.97	2	3.1^m	β^- γ_1 γ_2 γγ	1.79 0.58832 2.614	9* 13 2	22	aβγ s s	56D9 56B21 54M16 55W6	Other β^- groups. (56D9). Consistent decay scheme with energy matches.
Po $\xrightarrow{208}$ Bi 84 124 → 83 125 0^+ ?	> 0.85	1	2.9^y	ε γ_1 γ_2 $\gamma_1\gamma_2$	 0.285 0.270 ~0.60 0.570	 5* 5* 5* 5*		 scin scin scin scin	 55P7 55D6 55P7 55D6 55P7, 55D6	0.3γ/α ~ 3/10^5(55P7). ~ 7/10^3(55D6).
Rn $\xrightarrow{208}$ At 86 122 → 85 123 0^+	> 0		23^m	ε			80		55M24	
Tl $\xrightarrow{209}$ Pb 81 128 → 82 127 $1/2^+$ 5.3 $1/2^-$ $9/2^+$	3.12	3	2.2^m	γ_1 γ_2 γ_3	1.560 1.56 0.450 0.45 0.120 0.12	30* 3* 9* 1* 3* 1*		scin scin scin scin scin scin	56S20 55P8 56S20 55P8 56S20 55P8	Coincidence studies indicate all β to 2.23 level in Pb (56S20, 55P8)
Pb $\xrightarrow{209}$ Bi 82 127 → 83 126 $9/2^+$ 5.5 $9/2^-$	0.63	1	3.3^h	β_o^- no γ	0.63	1			55M25 55M25	F-K plot of β_o is linear (55M25).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
$_{84}Po_{125} \xrightarrow{209} {}_{83}Bi_{126}$ $1/2^-$ → $5/2^-$? $\underline{9/2^-}$	1.4	2	103^y	ϵ						All ϵ to 0.9 level (55P7).
				γ	0.910	18*		scin	55P7	
					0.865	17*		scin	55D6	
				$\epsilon_{LM}/\epsilon_K \sim 0.2$				scin	55P7	
$_{86}Rn_{123} \xrightarrow{209} {}_{85}At_{124}$	> 0		30^m	ϵ			83		55M24	
$_{81}Tl_{129} \xrightarrow{210} {}_{82}Pb_{128}$ $3, 4, 5^+$ 5.0 ? 2^+ 0^+	5.40	10	1.3^m	β^-	1.96	10		scin$\beta\gamma$	56M11	$\beta\gamma$, $\gamma\gamma$ studies suggest that all γ's are in cascade (56M11).
				γ_1	0.297	8		scin	56M11	
				γ_2	0.783	10		scin	56M11	
				γ_3	2.36	5		scin	56M11	
$_{82}Pb_{128} \xrightarrow{210} {}_{83}Bi_{127}$ 0^+ 5.4 0^- $\underline{1^-}$	0.063	1	19.4^y	β^-	0.017	2		scin	55L11	no γ with $0.025 < E\gamma < 0.032$. (56F4).
$_{83}Bi_{127} \xrightarrow{210} {}_{84}Po_{126}$ $\underline{1^-}$ 8.0 0^+	1.17 see comment	1	5.0^d see comment	β^-_0	1.17	1		s	55L12	non-unique shape does not permit good determination of β^-_0 end point energy. 2.6×10^{6y} Bi^{210} is possibly 25 kev below 5^d Bi^{210}.
					1.155	5		s	54P10	
$_{85}At_{125} \xrightarrow{210} {}_{84}Po_{126}$ 4, 5, 6 → 4, 5^- 0^+	4.1 Q < 4.3	1< 1	8.3^h	ϵ						Complex γ. Coincidence studies suggest decay scheme with highest level at 4.1 (54M17). Intensity of γ's suggest decay energy to 4.1 level ~ 0.1 (54M17).
				γ_1	0.2460	4*		s	54M17	
				γ_2	1.185	3*		s	54M17	
				$\gamma\gamma$, $\gamma\gamma(\theta)$					54M17	
$_{86}Rn_{124} \xrightarrow{210} {}_{85}At_{125}$ 0^+ → 4,5,6	> 0		2.7^h	ϵ			4		55M24	
$_{85}At_{126} \xrightarrow{211} {}_{84}Po_{127}$ 7/2, $9/2^-$ → $9/2^+$	> 0.67	1	7.2^h	ϵ						
				γ	0.671	10*		scin	54M17	
$_{86}Rn_{125} \xrightarrow{211} {}_{85}At_{126}$ $9/2^+$ → 7/2, $9/2^-$	> 1.34		16^h	ϵ			74		55M24	other γ's present (55M24).
				γ	1.34	2*		scin	55M24	

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	%	METHOD	REF.	COMMENTS
Pb -214-> Bi; 82 132 → 83 131; 0^+ 6.6 1^-	1.03	6	27^m	β^-_0	1.03	6	6	s	56D10	
Bi -214-> Po; 83 131 → 84 130; 1^- 7.8 0^+	3.21	3	20^m	β^-_0	3.26	3	19	s	56D11	Complex β's and γ's (56A2, 56D11, 56D12, 55J7, 55R10, 55R11, 55B25). Decay scheme, consistent with coincidence data, provides good energy matches (55J7).
					3.179	90	13	a	55R10	
				βγ, γγ		56D12, 55R10,			55R11	
Rn -221-> Fr; 86 135 → 87 134	> 0		25^m	β^-			80		56M12	
Fr -223-> Ra; 87 136 → 88 135; ≥ 6	≥ 1.3	2	21^m	β^-	1.3	2*	weak	s	54P11	Complex γ. Coincidence search suggests no γ coincidences (54H8).
					1.15	3*		scin	54H8	
				γ	0.310	30*		scin	54H8	
				γγ				scin	54H8	
Ra -225-> Ac; 88 137 → 89 136; $9/2^+$(?) ≥6.7 $7/2^-$?	0.35	1	14.8^d	β^-	0.32	3			55P8	Complex β, γ (55M25). (0.32β)(0.04γ) demonstrated.
					0.32	1		scin	55M25	
				γ	0.0395	20		scin	55P8	
					0.040	1		scin	55M25	
				βγ				55P8,	55M25	
Ac -226-> Th; 89 137 → 90 136; 1^-(?) ≥ 7.6 0^+	≥ 1.2	1	29^h	γ	0.159	5*		scin	55S11	
				γ	0.232	7*		scin	55S11	
Ac -227-> Th; 89 138 → 90 137; $3/2^-$ 6.7 $1/2^+$?	0.043	2	22^y	β^-_0	0.0455	10		pc	55B26	
				no γ, e^-, x				4πpc	55B26	
Pa -228-> Th; 91 137 → 90 138; 2^+; 0^+	> 1.72	2	22^h	ε, complex γ				s	5605	Information from both Pa-228->Th and Ac-228->Th is used in constructing level scheme (5605).
				Highest level reached	1.72	2*			5605	
Pa -230-> Th; 91 139 → 90 140; 0, 1^- ~7.7 0^+	≥ 1.4	1	17.7^d	β^+	~0.4	1*		s	5501	Complex γ, K/β^+ only ~2500 rather than > 2 x 10^4 estimated for Z = 90 (5501). See β^- branch below.

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF	COMMENTS
Pa $\xrightarrow{230}$ U 91 139 92 138 0, 1⁻ 6.4 0⁺	0.41	2	17.7^d	β_0^- no γ	0.405	20		s s	5501 5501	β^- branching ≤ 20 per cent (5501).
Th $\xrightarrow{231}$ Pa 90 141 91 140 7/2⁺ 5.9 5/2⁺ 3/2⁻	0.386	3	26^h	β^- γ_1 γ_2 βγ	0.305 0.0585 0.0255	10 3* 1*	40	s s s	56M13 56M13 56M13 56M13, 55S12	Complex γ. (56M13). Decay scheme with good γ energy sum matches. (56M13). Suggest 0.3β to 0.084 level in Pa^{231} (56M13).
Pa $\xrightarrow{232}$ U 91 141 92 140 9.20 0⁺	1.25	2	1.31^d	β_0^-	1.27	3*	2	s	54B10	Lower energy β^- and γ present (54B10).
Pa $\xrightarrow{233}$ U 91 142 92 141 8.9 5/2⁺	0.568	4	27.0^d	β_0^-	0.568 0.566	5* 6*	5 5	s s	5502 54B11	Complex lower energy β^- and γ. (5502, 54B11). Decay scheme has good energy matches (5502, 54B11).
Th $\xrightarrow{234}$ Pa 90 144 91 143 0⁺ 6.6 1.2ᵐ 6.7ʰ	≥ 0.192	1	24^d	β^- γ βγ, γγ	0.100 0.10 0.0914 0.093	2* 1 5* 2	35	sβγ scinβγ s scin	55D7 54J5 5606 54J5 56V3, 5606, 55D7, 54J5	Data indicates that I. T. energy is very small (5606, 54J5).
Pa $\xrightarrow{234}$ U 91 143 92 142 1.2ᵐ 6.7ʰ 7.3 ? 0⁺	2.28	6	6.66^h	β^-	1.13	5	13	s	5606	Complex γ(5606, 55D7, 54J5). Decay scheme by (5606) has 1.13β to level at 1.145.
Np $\xrightarrow{234}$ U 93 141 92 142 8.3 0⁺	1.8	1	4.4^d	β_0^+ β^+/ϵ = 5 x 10^{-4} γ	~0.8 1.565	1* 30*		a,s s scin	55P9 55P9 56H12	Complex γ (56H12). β^+/ϵ ratio indicates that β_0^+ is probably g. s. transition. (56H12).
Np $\xrightarrow{235}$ U 93 142 92 143 5/2⁻ 6.5 5/2⁺	0.13	2	410^d	ε ϵ_L/ϵ_K = 30 no $U^{235}\gamma$		 2		 pc, scin scin	 56H13 56H13	ϵ_L/ϵ_K = 30 implies Σ_ϵ = 0.13(56H13)
Np $\xrightarrow{236}$ Pu 93 143 94 142 2⁺ 6.6 0⁺	0.515	7	22^h see comment	β_0^-	0.518 0.53	10 3	60	s	5609 S54P11	β_0^- reported to have non-linear FK plot (5609). 22^h Np^{236} may not be g. s. since long lived (> 5000^y) Np^{236} has been found in U^{238}(d, 4n) bombardment (55S13).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q		HALF-LIFE	DECAY DATA						COMMENTS
	MEV	ERROR		TYPE	MEV	ERROR	%	METHOD	REF.	
Pa -237-> U 91 146 92 145	> 0		10^m							p. 6.7^d U^{237} (54C8).
U -237-> Np 92 135 93 144 1/2 6.1 5/2⁻	0.514	3	6.75^d	β^-	0.249	5	74	s	56B22	Decay scheme established with the aid of information from Am^{241} decay. Complex γ. 0.25β and 0.27γ considered to be in cascade (56B22).
				γ	0.268	2*		s	56B22	
Np -238-> Pu 93 145 94 144 2^+ 8.5 2^+ 0^+	1.305	8	2.1^d	β^-	1.25	2*		s	55R12	(1.25β)(0.04γ) demonstrated. (55R12).
				γ	0.0440	2*		s	55R12	
				βγ, γγ					55R13	
Np -239-> Pu 93 146 94 145 1/2 >7.2 $1/2^+$	0.719 or 0.729 see comment	6 6	2.35^d	β^-	0.343	15		scinβγ	55E2	Complex β's and γ's (56H14, 55L13, 55E2). Decay scheme, consistent with predictions of collective model, displaces earlier level schemes upward by 7.8 kev and predicts a first excited state at 0.00785. (56H14, 56H15). However, level scheme suggests spin of Np^{239} is 5/2 or 3/2 and that $0.7\beta^-$ may not go to Pu^{239} g. s.
Np -240-> Pu 93 147 94 146 0, 1 ≥5.8 0^+	2.05	6	1^h	β^-	0.90	5*		s	54L7	Other γ's present (56L10, 54L7). 7.3^m Np^{240} is probably isomer of 1^h Np^{240} (54L7).
				γ	1.14 1.16	3* 3*		scin scin	56L10 54L7	
				γγ, βγ					54L7	
Am -240-> Pu 95 145 94 146 0^+	> 1.40	3	47^h	ε						
				γ	1.40	3		scin	54G3	Lower energy γ's present (54G3).
Pu -241-> Am 94 147 95 146 5/2 5.6 5/2⁻	0.0208	2	12.9^y	β_o^-	0.0208 0.0205	2 10*		s s	56S21 54P11	F-K plot of β_o has allowed shape (56S21). α branching ~ 4×10^{-5} (55J8).
				no γ				s	56S21	
Cm -241-> Am 96 145 95 146 5/2⁻	> 0.592	10	35^d	γ	0.592	10		scin	54G3	0.47γ present (54G3).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF	COMMENTS
$_{95}Am_{147} \xrightarrow{242} {}_{94}Pu_{148}$ 16^h 0^- 100^y $4^+, 5^\pm$ 2^+ 0^+	> 0.0445	3	100^y	ϵ γ	 0.0445 0.0446	 4* 3		 s s	 55B27 55H13	ϵ branching ~ 10 percent. (55H12).
$_{95}Am_{147} \xrightarrow{242} {}_{96}Cm_{146}$ 16^h 0^- 100^y $4^+, 5^\pm$ 11.3 2^+ 0^+	0.633	5	100^y	β^- γ	0.585 0.0422 0.0422	10 4* 3		s s s	55H12 55B27 55H13	ΔJ = 2; yes shape observed for β^- decay to 2^+ state in Cm^{242} implies spin 0^- for 16^h Am^{242}. Hence 100^y Am^{242} probably has spin 4, 5 and 0.59β is probably to 2^+ state in Cm^{242}(55B27). However, (55H12) suggests β branching to 2^+ <u>and</u> 0^+ g. s.
$_{95}Am_{149} \xrightarrow{244} {}_{94}Pu_{150}$ 0^- 0^+	> 0		26^m	ϵ				chem	55F8	ϵ/β^- = 3.9 x 10^{-4} (55F8).
$_{95}Am_{149} \xrightarrow{244} {}_{96}Cm_{148}$ 0^- 6.1 0^+	1.5	1	26^m	β^-_0 no γ	1.5	1		scin scin	54G4 54G4	
$_{94}Pu_{151} \xrightarrow{245} {}_{95}Am_{150}$	> 0		10.5^h	p p	2^h Am^{245} 25^m Am^{245}			chem chem chem	56B23 55F9 55B28	
$_{95}Am_{150} \xrightarrow{245} {}_{96}Cm_{149}$ ≥ 6.2	≥ 0.905	5	2.03^h	β^- $\gamma\gamma$	~0.86 0.905	5* 5		a s scin	55F9 55B28 55B28	Complex γ (55B28, 55F9).
$_{97}Bk_{148} \xrightarrow{245} {}_{96}Cm_{149}$ 6.7	0.66	3	5^d	γ_1 γ_2 $\gamma_1 \gamma_2$	0.250 0.380(double?)	2*		s,scin s, scin	56M14 56M14 56M14	ϵ_L/ϵ_K to 0.25 level = 0.33, to ~ 0.39 level = 5 to 10 (56M14). Estimate Q from $\gamma\gamma$ and ϵ_L/ϵ_K.
$_{98}Cf_{147} \xrightarrow{245} {}_{97}Bk_{148}$	> 0		45^m	ϵ			70	chem	56C7	p 5^d Bk (56C7).

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	MEV	ERROR	%	METHOD	REF.	COMMENTS
Pu $\xrightarrow{246}$ Am 94 152 95 151 0^+ 7.4 1^- ?	0.375	30	11^d	β_1^-	0.33	3	27	scin$\beta\gamma$	56H16	$(0.33\beta^-)(0.05\gamma)$ (56H16).
					0.15	1	73	scin$\beta\gamma$	56H16	$(0.15\beta^-)(0.175\gamma)$ gives good energy match. (56H16).
				β_2^-	0.15	1		a	55E3	
				γ	0.047	2*		scin	56H16	
					0.043	2		scin	55E3	
				$\beta\gamma$, $\gamma\gamma$				scin	56H16	
Am $\xrightarrow{246}$ Cm 95 151 96 150 ? 0^+	2.29	5	25^m	β^-	1.222	4*	strong	scin$\beta\gamma$	55E3	$(1.2\beta^-)(1.07\gamma)$ reported by (55E3).
				β^-	>1.2		weak	scin	55E3	
				γ	1.069	20*		scin	55E3	
				$\beta\gamma$, $\gamma\gamma$					55E3	
Bk $\xrightarrow{246}$ Cm 97 149 96 150 0^+	> 0.82	2	1.8^d	ϵ (K, L x rays)				pc,scin	54H9	
				γ	0.82	2*	40	scin	54H9	
E $\xrightarrow{246}$ Cf 99 147 98 148 0^+	> 0		$?^m$	ϵ, p. 1.5^d Cf				chem	54G4	
Cf $\xrightarrow{247}$ Bk 98 149 97 150	> 0.13		2.5^h	ϵ (K, L x rays observed)				pc,scin	54H9	
E $\xrightarrow{248}$ Cf 99 149 98 150 0^+	> 0		25^m	ϵ				chem	56C7	
Cm $\xrightarrow{249}$ Bk 96 153 97 152 6.0	≥ 0.9	1	65^m	β^-	0.9	1*		a(?)	56F5	
Bk $\xrightarrow{249}$ Cm 97 152 98 151 ≥ 6	≥ 0.08	2	290^d	β^-	0.08	2		a	54M18	a branching ~ 10^{-3} per cent (54M18)
E $\xrightarrow{249}$ Cf 99 150 98 151	> 0		2^h	ϵ				chem	56H17	ϵ/a ~ 760 (56H17)
Bk $\xrightarrow{250}$ Cm 97 153 98 154 1^- 2^+ ≥7.5 0^+	1.9	1	3.1^h	β_0^-	1.9	1*		scin	54G4	$(0.9\beta)(\sim0.9\gamma)$ provides energy match (54G4).
E $\xrightarrow{250}$ Cf 99 151 98 152	> 0		8^h	ϵ					56H17	
E $\xrightarrow{251}$ Cf 99 152 98 153	> 0		1.5^d	ϵ					56H17	ϵ/a ~ 190 (56H17)

TABLE I (*continued*).

DISINTEGRATION	ADOPTED Q MEV	ADOPTED Q ERROR	HALF-LIFE	DECAY DATA TYPE	DECAY DATA MEV	DECAY DATA ERROR	DECAY DATA %	DECAY DATA METHOD	DECAY DATA REF.	COMMENTS
Cf (98 155) $\xrightarrow{253}$ E (99 154)	> 0		18^d	β^-					54M18	
									54C9	
Fm (100 153) $\xrightarrow{253}$ E (99 154)	> 0		4.5	ε			89	chem	57A2	
E (99 155) ----> Fm (100 154)	< 1.70	5	> 2y	From 38^h state						+ photons / $100\beta_1^-$ (56J9).
38^h				β^-	1.04	4		a	56J9	
					1.1	1		scin	54C9	
> 2y 2^+				γ	0.660	15	70^+	scin	56J9	
0^+										
E (99 156) $\xrightarrow{255}$ Fm^{255} (100 155)	> 0		27^d	p	22^h Fm^{255}				56J9	
									54C9	

(References to Table I appear on pages 826–830.)

REFERENCES

1948

48S1 F.B.Shull, Phys. Rev. 74, 917.

1949

49D1 R.B.Duffield, J.D.Knight, Phys. Rev. 76, 573.

1951

51H1 E.K.Hyde, G.D.O'Kelley, Phys. Rev. 82, 944.
51K1 H.B.Keller, J.M.Cork, Phys. Rev. 84, 1079.

1952

52B1 F.D.S.Butement, Proc. Phys. Soc. 65A, 254.

1953

53B1 Bendel, Shore, Brown, Becker, Phys. Rev. 90, 888.
53C1 Cork, Nester, LeBlanc, Brice, Phys. Rev. 92, 119.
53F1 R.W.Fink, Thesis, Univ. of Rochester (1953); NSA 9, 24B (1955).
53K1 L.Katz, R.G.Baker, R.Montalbetti, Can. J. Phys. 31, 250.
53L1 L.J.Lidofsky, unpublished data.
53M1 J.P.Mize, D.J.Zaffarano, ISC 347.
53M2 K.Murakawa, T.Kamei, Phys. Rev. 92, 235.
53N1 R.H.Nussbaum, R.van Lieshout, A.H.Wapstra, Phys. Rev. 92, 207.
53S1 A.B.Smith, Dissertation Abstr. 13, 849; NSA 9, 24B (1955).
53W1 C.S.Wu, A.Schwarzschild, Phys. Rev. 91, 483A; private communication (March 1957).

1954

54A1 D.P.Ames, M.E.Bunker, J.W.Starner, private communication to TID 5300 (54A44).
54A2 A.H.W.Aten,Jr., J.Manassen, G.D.de Feyfer, Physica 20, 665.
54A3 J.A.Arnold, Phys. Rev. 93, 743.
54A4 D.E.Alburger, M.H.L.Pryce, Phys. Rev. 95, 1482.

54B1 G.A.Bazorgan, J.W.Irvine, C.D.Coryell, Phys. Rev. 95, 781.
54B2 G.Bolla, S.Terrani, S.Zappa, Nuovo Cim. 12, 875.
54B3 S.B.Burson, W.C.Jordan, J.M.LeBlanc, Phys. Rev. 96, 1555.
54B4 M.E.Bunker, J.P.Mize, J.W.Starner, Phys. Rev. 94, 1694.
54B5 Bertolini, Bisi, Lazzarini, Zappa, Nuovo Cim. 11, 539.
54B6 F.I.Boley, Phys. Rev. 94, 1078.
54B7 Bashilov, Anton'eva, Blinov, Dzhelepov, Izvest. Akad. Nauk Ser. FIZ SSSR 18, 43; NSA 9, 24B (1955).
54B8 H.N.Brown, R.A.Becker, Phys. Rev. 96, 1372.
54B9 A.Bisi, L.Zappa, Nuovo Cim. 12, 539.
54B10 Browne, Hoffman, Smith, Bunker, Mize, Starner, Moore, Balagna, Phys. Rev. 96, 827A.
54B11 W.D.Brodie, Proc. Phys. Soc. 67A, 397.

54C1 R.S.Caird, A.C.G.Mitchell, Phys. Rev. 94, 412.
54C2 Cohen, Charpie, Handley, Olson, Phys. Rev. 94, 953.
54C3 Cork, Brice, Hickman, Schmid, Phys. Rev. 93, 1059.
54C4 J.M.Cork, M.K.Brice, L.C.Schmid, Phys. Rev. 96, 1259.
54C5 E.L.Church, M.Goldhaber, Phys. Rev. 95, 626A.
54C6 Cork, Brice, Schmid, Hickman, Nine, Phys. Rev. 94, 1218.
54C7 Coryell, Herber, Sugihara, Bennett, Phys. Rev. 95, 298A.
54C8 W.W.T.Crane, G.M.Iddings, Phys. Rev. 95, 1702.
54C9 Choppin, Thompson, Ghiorso, Harvey, Phys. Rev. 94, 1080.

54D1 H.Daniel, W.Bothe, Z. Naturf. 9a, 402.
54D2 R.M.Diamond, Phys. Rev. 95, 410.
54D3 DeRaad, Middelkoop, B.Van Nooyen, P.M.Endt, Physica 20, 1278.
54D4 D.Dixon, A. McNair, S.C.Curran, Phil. Mag. 45, 683.
54D5 D.Dixon, A.McNair, Phil. Mag. 45, 1099.

54E1 B.Elbek, K.O.Nielsen, O.B.Nielsen, Phys. Rev. 95, 96.

54F1 H.H.Forster, A.Oswald, Phys. Rev. 96, 1030.
54F2 L. Feuvrais, T.Yuasa, Compt.rend. 239, 1627.
54F3 J.Flinta, S.Eklund, Arkiv. Fysik 7, 401.
54F4 H.L.Finston, W.Bernstein, Phys. Rev. 96, 71.
54F5 R.C.Fix, C.D.Coryell, MIT LSNE Prog. Rep., p. 37, May (1954); NSA 9, 24B (1955).
54F6 V.K.Fischer, Phys. Rev. 96, 1549.
54F7 Fowler, Kruse, Keshishian, Klotz, Mellor, Phys. Rev. 94, 1082.

54G1 J.B.Gerhart, Phys. Rev. 95, 288.
54G2 Grace, Johnson, Kurti, Lemmer, Robinson, Phil. Mag. 45, 1192.
54G3 R.A.Glass, UCRL-2560.
54G4 Ghiorso, Thompson, Choppin, Harvey, Phys. Rev. 94, 1081.
54G5 Ghiorso, Rossi, Harvey, Thompson, Phys. Rev. 93, 257.

54H1 E.K.Hyde, M.G.Florence, F.S.Stephens, UCRL-2813.
54H2 B.C.Haldar, E.O.Wiig, Phys. Rev. 94, 1713.
54H3 T.H.Handley, E.L.Olson, Phys. Rev. 96, 1003.
54H4 T.H.Handley, E.L.Olson, Phys. Rev. 93, 524.
54H5 T.H.Handley, Phys. Rev. 94, 945.
54H6 T.H.Handley, E.L.Olson, Phys. Rev. 94, 968.
54H7 J.R.Huizenga, C.M.Stevens, Phys. Rev. 96, 548.
54H8 E.K.Hyde, Phys. Rev. 94, 1221.
54H9 E.K.Hulet, S.G.Thompson, A.Ghiorso, Phys. Rev. 95, 1702.

54J1 Jones, Phillips, Johnson, Wilkinson, Phys. Rev. 96, 547.
54J2 H.Jaffe, UCRL-2537.
54J3 M.C.Joshi, B.V.Thosar, Phys. Rev. 96, 1022.
54J4 M.W.Johns, S.V.Nablo, Phys. Rev. 96, 1599.
54J5 S.A.E.Johansson, Phys. Rev. 96, 1075.

54K1 Koerts, Schwarzschild, Gold, Wu, Phys. Rev. 95, 612A.
54K2 J.D.Kurbatov, B.B.Murray, M.Sakai, Phys. Rev. 98, 674.
54K3 D.B.Kochendorfer, D.J.Farmer, Phys. Rev. 96, 855A.
54K4 M.I.Kalkstein, J.M.Hollander, Phys. Rev. 96, 730.
54K5 Keshnian, Kruse, Klotz, Fowler, Phys. Rev. 96, 1050.
54K6 L.Koerts, Phys. Rev. 95, 1358.
54K7 J.C.Knight, T.H.Braid, H.O.W.Richardson, Proc. Phys. Soc. 67A, 881.

54L1 T.Lindqvist, A.C.G.Mitchell, Phys. Rev. 95, 1535.
54L2 D.C.Lu, W.H.Kelley, M.L.Wiedenbeck, Phys. Rev. 95, 1533.
54L3 C.Levi, L.Papineau, Compt.rend. 238, 2313.
54L4 C.Levi, L.Papineau, Compt.rend. 239, 1782.
54L5 N.H.Lazar, Phys. Rev. 95, 292.
54L6 D.C.Lu, M.L.Wiedenbeck, Phys. Rev. 94, 501.
54L7 R.M.Lessler, UCRL-2531, p. 50 (1954).

54M1 A.Moljk, S.C.Curran, Phys. Rev. 96, 395.
54M2 Th.Mayer-Kuckuk, Z. Naturf. 9a, 338.
54M3 Maeder, Wapstra, Nijgh, Ornstein, Physica 20, 521.
54M4 L.Madansky, F.Rasetti, Phys. Rev. 94, 412.
54M5 W.C.Middelkoop, B.Van Nooyen, P.M.Endt, private communication to TID-5300 (54M76).
54M6 P.Macq, Ann. Soc. Sci. Bruxelles 67, 30[illegible]; NSA 8 24B (1954).
54M7 L.Marquez, Phys. Rev. 95, 67.
54M8 J.Moreau, Compt. rend. 239, 1130.
54M9 H.B.Mathur, E.K.Hyde, Phys. Rev. 96, 126.
54M10 J.Moreau, Compt. rend. 239, 820.
54M11 J.Moreau, Arkiv Fysik 7, 391.
54M12 H.B.Mathur, E.K.Hyde, Phys. Rev. 95, 708.
54M13 A.C.G.Mitchell, E.Hebb, Phys. Rev. 95, 727.
54M14 K.Murakawa, Phys. Rev. 96, 1543.
54M15 K.Murakawa, Phys. Rev. 98, 1285.
54M16 D.I.Meyer, F.M.Schmidt, Phys. Rev. 94, 927.
54M17 J.W.Mihelich, A.W.Schardt, E.Segre, Phys. Rev. 95, 1508.

54M18 Magnusson, Studier, Fields, Stevens, Mech, Friedman, Diamond, Huizenga, Phys. Rev. 96, 1576; 94, 1083.

54N1 M.E.Nahmias, J. Phys. radium 15, 568.
54N2 M.E.Nahmias, A.H.Wapstra, J. Phys. radium 15, 570.
54N3 R.T.Nichols, E.N.Jensen, Phys. Rev. 94, 369.
54N4 Nussbaum, Van Lieshout, Wapstra, Verster, Ten Haaf, Nijgh, Ornstein, Physica 20, 555.
54N5 R.H.Nussbaum, A.H.Wapstra, M.J.Sterk, private communication to TID-5300 (54N23).
54N6 Nussbaum, Wapstra, Van Lieshout, Nijgh, Ornstein, Physica 20, 571.
54N7 R.H.Nussbaum, R. Van Lieshout, Physica 20, 440.

54O1 Ong Ping Hok, P.Kramer, G.Meijer, Physica 20, 1200.
54O2 J.L.Olsen, G.D.O'Kelley, Phys. Rev. 95, 1539.

54P1 J.R.Penning, F.H.Schmidt, Phys. Rev. 94, 779A.
54P2 F.T.Porter, private communication to TID-5300 (54P05).
54P3 Pohm, Lewis, Talboy, Jensen, Phys. Rev. 95, 1523.
54P4 R.Patterson, Phys. Rev. 95, 303A.
54P5 M.L.Perlman, J.Welker, Phys. Rev. 95, 133, 613A.
54P6 C.H.Pruett, R.G.Wilkinson, Phys. Rev. 96, 1340.
54P7 R.W.Pringle, W.Turchinetz, H.W.Taylor, Phys. Rev. 95, 115.
54P8 T.O.Passell, M.C.Michel, I.Bergstrom, Phys. Rev. 95, 999.
54P9 J.R.Prescott, Proc. Phys. Soc. 67A, 540.
54P10 E.A.Plassmann, L.M.Langer, Phys. Rev. 96, 1593.
54P11 T.O.Passell, UCRL-2528.

54S1 Simanton, Rightmire, Long, Kohman, Phys. Rev. 96, 1711.
54S2 Sheline, Johnson, Bell, Davis, McGowan, Phys. Rev. 94, 1642.
54S3 S.E.Singer, W.S.Emmerich, J.D.Kurbatov, Phys. Rev. 94, 113.
54S4 H.Selig, NYO-6626; NSA 10, 24.
54S5 R.E.Slattery, D.C.Lu, M.L.Wiedenbeck, Phys. Rev. 96, 465.
54S6 A.D.Suttle, Jr., W.F.Libby, Phys. Rev. 95, 866.

54T1 S.Thulin, J.Moreau, H.Atterling, Arkiv Fysik 8, 219.
54T2 S.Thulin, J.Moreau, H.Atterling, Arkiv Fysik 8, 229.
54T3 S.Thulin, K.Nybo, Arkiv Fysik 7, 289.

54V1 J.Varma, C.E.Mandeville, Phys. Rev. 94, 91.
54V2 J.Verhaega, J.Demuynck, Compt. rend. 239, 1374.

54W1 Calvin Wong, Phys. Rev. 95, 765.
54W2 K.Way, M.Wood, Phys. Rev. 94, 119.
54W3 Wapstra, Maeder, Nijgh, Ornstein, Physica 20, 169.
54W4 A.H.Wapstra, Arkiv Fysik 7, 279.

1955

55A1 D.E.Alburger, A.W.Sunyar, Phys. Rev. 99, 695.
55A2 P.Axel, J.D.Fox, R.H.Parker, Phys. Rev. 97, 975.
55A3 A.H.W.Aten, Jr., de Vries-Hamerling, Physica 21, 597.
55A4 P.Avignon, A.Michalowicz, R.Bouchez, J. Phys. radium 16, 404.
55A5 D.E.Alburger, B.J.Toppel, Phys. Rev. 100, 1357.
55A6 T.Azuma, J. Phys. Soc. Japan 10, 167; NSA 9 24B.
55A7 B.P.Adyasevich, B.D.Groshev, A.M.Demidov, Conf. Acad. Sci. USSR on Peaceful Use of Atomic Energy, Phys. Math. Sci. p. 270, July (1955). Consultants Bureau Trans. p. 150; NSA 10 24B (1956).
55A8 A.H.W.Aten, Jr., G.D.de Feyfer, Physica 21, 543.
55A9 Aten, Jr., De Feyfer, Sterk, Wapstra, Physica 21, 740.
55A10 Andersson, Arbman, Bergstrom, Wapstra, Phil. Mag. 46, 70.

55B1 J.W.Blue, E.Bleuler, Phys. Rev. 100, 1324.
55B2 M.E.Bunker, J.W.Starner, Phys. Rev. 97, 1272.
55B3 A.Bisi, E.Germagnoli, L.Zappa, Nuovo Cim. 2, 1052.
55B4 F.I.Boley, private communication to TID-5300 (55B36).
55B5 J.Bellicard, A.Moussa, Compt. rend. 241, 1202.
55B6 F.D.S.Butement, E.G.Prout, Phil. Mag. 46, 357.
55B7 A.Bisi, E.Germagnoli, L.Zappa, Nuovo Cim. 2, 290.
55B8 A.Bisi, S.Terrani, L.Zappa, Nuovo Cim. 2, 1297.
55B9 D.Bes, H.Bosch, Compt. rend. 240, 294.
55B10 M.E.Bunker, J.P.Mize, J.W.Starner, Phys. Rev. 99, 659A.
55B11 G.B.Baro, P.Rey, W.Seelmann-Eggebert, Z. Naturf. 10a, 81.
55B12 S.B.Burson, J.M.LeBlanc, D.W.Martin, Phys. Rev. 99, 660A.
55B13 G.Bertolini, M.Bettoni, E.Lazzarini, Nuovo Cim. 2, 273; 1, 746.
55B14 A.R.Brosi, B.H.Ketelle, Phys. Rev. 100, 169.
55B15 Bolotin, Pruett, Roggencamp, Wilkinson, Phys. Rev. 99, 62.
55B16 G.R.Bishop, J.P.Perez y Jorba, Phys. Rev. 98, 89.
55B17 R.Barloutaud, R.Ballini, Compt. rend. 241, 389.
55B18 A.Bisi, S.Terrani, L.Zappa, Nuovo Cim. 2, 172.
55B19 F.D.S.Butement, A.J.Poe, Phil. Mag. 46, 482.
55B20 A.Bisi, S.Terrani, L.Zappa, Nuovo Cim. 1, 291.
55B21 Baggerly, Marmier, Boehm, DuMond, Phys. Rev. 100, 1364.
55B22 H.Brysk, M.E.Rose, ORNL 1830.
55B23 Brunner, Guhl, Halter, Leisi, Helv. Phys. Acta 28, 475A.
55B24 Bergkvist, Bergstrom, Herrlander, Hultberg, Slatis, Sokolowski, Wapstra, Wiedling, Phil. Mag. 46, 65.
55B25 G.Backenstoss, K.Wohlleben, Z.Naturf. 10a, 384.
55B26 W.Beckmann, Z. Phys. 142, 585.
55B27 S.A.Baranov, K.N.Shliagin, Conf. Acad. Sci. USSR on Peaceful Use of Atomic Energy, Phys. Math. Sci. p. 251, July (1955); Consultants Bureau Trans. p. 183.
55B28 Browne, Hoffman, Crane, Balagna, Higgins, Barnes, Hoff, Smith, Mize, Bunker, J. Inorg. Nucl. Chem. 1, 254; NSA 10, 24B, p. 126 (1956).

55C1 S.G.Cohen, S.Ofer, Phys. Rev. 100, 856.
55C2 B.Craseman, D.L.Manley, Phys. Rev. 98, 66.
55C3 J.M.Cork, M.K.Brice, L.C.Schmid, Phys. Rev. 99, 703.
55C4 J.B.Cumming, Phys. Rev. 99, 1645A.
55C5 W.A.Cassatt, Jr., W.W.Meinke, Phys. Rev. 99, 760.
55C6 Cork, Brice, Schmid, Helmer, Phys. Rev. 100, 188.
55C7 W.A.Cassatt, Jr., W.W.Meinke, Phys. Rev. 100, 1372.
55C8 O.F.Coleman, Phil. Mag. 46, 1132.

55D1 E.F.de Haan, G.J.Sizoo, Physica 21, 818.
55D2 R.W.P.Drever, A.Moljk, Phil. Mag. 46, 1337.
55D3 Dick, Foucher, Perrin, Vartapetian, Compt. rend. 240, 1335.
55D4 M.C.Day, Jr., G.W.Eakins, A.F.Voigt, Phys. Rev. 100, 796.
55D5 H.de Waard, Phil. Mag. 46, 445.
55D6 E.H.Daggett, G.R.Grove, Phys. Rev. 99, 1.
55D7 E.F.de Haan, G.J.Sizoo, P.Kramer, Physica, 21, 803.

55E1 Eberhardt, Geiss, Lang, Herr, Merz, Z. Naturf. 10a, 796.
55E2 D.Engelkemeir, L.B.Magnusson, Phys. Rev. 99, 135.
55E3 Engelkemeir, Fields, Fried, Pyle, Stevens, Asprey, Browne, Smith, Spence, J. Inorg. Nucl. Chem. 1, 345; NSA 10, 24B, p. 126 (1956).

55F1 R.L.Folger, P.C.Stevenson, G.T.Seaborg, Phys. Rev. 98, 107.
55F2 D.J.Farmer, Phys. Rev. 99, 659A.
55F3 H.H.Forster, A.Rosen, Nuovo Cim. 1, 972.
55F4 Farrelly, Koerts, Benczer, van Lieshout, Wu, Phys. Rev. 99, 1440.
55F5 I.Franz, J.Rodriguez, H.Carminatti, Z. Naturf. 10a, 82.
55F6 H.H.Forster, J.S.Wiggins, Nuovo Cim. 2, 854.
55F7 V.K.Fischer, Phys. Rev. 99, 764.
55F8 Fields, Gindler, Harkness, Studier, Huizenga, Friedman, Phys. Rev. 100, 172.
55F9 Fields, Studier, Friedman, Diamond, Sjoblom, Sellers, J. Inorg. Nucl. Chem. 1, 254; NSA 10, 24B, p. 126 (1956).

55G1 N.W.Glass, J.R.Richardson, Phys. Rev. 98, 1251.
55G2 W.E.Graves, A.C.G.Mitchell, Phys. Rev. 97, 1033.
55G3 M.J.Glaubmann, Phys. Rev. 98, 645.
55G4 R.L.Graham, J.L.Wolfson, M.A.Clark, Phys. Rev. 98, 1173A.
55G5 S.F. Grard, L.Danguy, J.Franeau, J. phys. radium 16, 839.

55H1 T.H.Handley, W.S.Lyon, Phys. Rev. 99, 755.
55H2 K.P.Howard, T.A.Pond, P.S.Jastram, private communication to TID-5300 (55H24).
55H3 G.Herrmann, F.Strassmann, Z. Naturf, 10a, 146.
55H4 R.W.Hayward, D.D.Hoppes, H.Ernst, Phys. Rev. 98, 231A.
55H5 J.M.Hollander, M.I.Kalkstein, Phys. Rev. 98, 260A.
55H6 E.Hebb, Phys. Rev. 97, 987.
55H7 H.S.Hans, B.Saraf, C.E.Mandeville, Phys. Rev. 97, 1267.

55H8 T.H.Handley, W.S.Lyon, Phys. Rev. 99, 1415.
55H9 T.H.Handley, W.S.Lyon, E.L.Olson, Phys. Rev. 98, 688.
55H10 E.N.Hatch, F.Boehm, J.W.M.DuMond, private communication to NSA 10, 24B, p. 92 (1956).
55H11 S.K.Haynes, W.T.Achor, J. phys. radium 16, 48.
55H12 Hoff, Jaffe, Passell, Stephens, Hulet, Thompson, Phys. Rev. 100, 1403.
55H13 Ref. 55H12; E.L.Church, ibid.

55J1 M.W.Johns, B.C.Chidley, I.R.Williams, Phys. Rev. 99, 1645A.
55J2 O.E.Johnson, R.G.Johnson, L.M.Langer, Phys. Rev. 98, 1517.
55J3 F.A.Johnson, Can. J. Phys. 33, 841.
55J4 J.T.Jones,Jr., E.N.Jensen, Phys. Rev. 97, 1031.
55J5 S.E.Johansson, Phys. Rev. 100, 835.
55J6 Joly, Brunner, Halter, Huber, Helv. Phys. Acta 28, 403.
55J7 S.E.Johansson, Arkiv Fysik 9, 561.
55J8 Jones, Schuman, Butler, Cowper, Eastwood, Jackson, KAPL-1378.

55K1 O.C.Kistner, B.M.Rustad, O.Kofoed-Hansen, private communication (March 1957).
55K2 O.Kofoed-Hansen, A.Nielsen, Kgl. Danske Videnskab. Selskab Mat.-Fys. Medd. 29, No. 15; Phys. Rev. 96, 1045 (1954).
55K3 S.Kono, J. Phys. Soc. Japan, 10, 495; NSA 9 24B (1955).
55K4 J.C.Kluyver, C.Van der Leun, Physica 21, 604.
55K5 L.E.Killion, Dissertation Abstr. 15, 858; NSA 9 24B (1955).
55K6 B.Kahn, W.S.Lyon, Phys. Rev. 98, 58.
55K7 S.Katcoff, Phys. Rev. 99, 1618.
55K8 Koerts, Macklin, Farrelly, van Lieshout, Wu, Phys. Rev. 98, 1230.
55K9 G.L.Keister, E.B.Lee, F.H.Schmidt, Phys. Rev. 97, 451.
55K10 Kreger, McIsaac, Mackin, Lai, Phys. Rev. 100, 953.
55K11 J.D.Knight, E.W.Baker, Phys. Rev. 100, 1334.

55L1 M.J.Laubitz, Proc. Phys. Soc. 68A, 1033.
55L2 T.Lindqvist, C.S.Wu, Phys. Rev. 100, 145.
55L3 W.S.Lyon, T.H.Handley, Phys. Rev. 100, 1280.
55L4 W.S.Lyon, B.Kahn, Phys. Rev. 99, 728.
55L5 H.R.Lemmer, O.J.A.Segaert, M.A.Grace, Proc. Phys. Soc. 68A, 701.
55L6 L.Linder, G.A.Brinkman, A.C.Pieterse, Physica 21, 745.
55L7 J.M.LeBlanc, J.M.Cork, S.B.Burson, Phys. Rev. 97, 750.
55L8 D.C.Lu, W.H.Kelly, M.L.Wiedenbeck, Phys. Rev. 97, 139.
55L9 J.Laberrigue-Frolow, M.Lederer, J. Phys. radium 16, 346.
55L10 N.H.Lazar, E.D.Klema, Phys. Rev. 98, 710.
55L11 G.M.Lewis, Proc. Phys. Soc. 68A, 735L.
55L12 Lidofsky, Benczer, Macklin, Wu, Phys. Rev. 98, 1186A.
55L13 H.W.Lefevre, E.M.Kinderman, H.H.Van Tuyl, Phys. Rev. 100, 1374.

55M1 H.Morinaga, Phys. Rev. 100, 431.
55M2 M.McKeown, G.Scharff-Goldhaber, private communication to TID-5300 (55M47).
55M3 H.Morinaga, E.Bleuler, Phys. Rev. 1236A.
55M4 Th.Mayer-Kuckuk, H.Daniel, Z. Naturf. 10a, 168.
55M5 Middelkoop, Heyligers, Rietjens, Van den Bold, Endt, Physica 21, 897.
55M6 L.Madansky, F.Rasetti, Phys. Rev. 97, 837.
55M7 Mathur, Hyde, Levine, Kofstad, Phys. Rev. 97, 117.
55M8 H.B.Mathur, E.K.Hyde, Phys. Rev. 98, 79.
55M9 K.Murakawa, Phys. Rev. 100, 1369.
55M10 H.Mark, C.McClelland, C.Goodman, Phys. Rev. 98, 1245.
55M11 S.W.Mead, M.D.Petroff, W.O.Doggett, Phys. Rev. 100, 1794A.
55M12 C.L.McGinnis, Phys. Rev. 97, 93.
55M13 Murin, Preobrazhensky, Yutlandov, Yakimov, Conf. Acad. Sci. on Peaceful Use of Atomic Energy, Chem. Sci. p. 160, July 1955; NSA 10, 24B (1956).
55M14 Mallmann, Aten,Jr., Bess, deMcMillan, Phys. Rev. 99, 7.
55M15 N.Marty, J. Phys. radium 16, 458.
55M16 N.Marty, Compt. rend. 241, 385.
55M17 J.C.D.Milton, J.S.Fraser, G.M.Milton, Phys. Rev. 98, 1173A.
55M18 J.P.Mize, M.E.Bunker, J.W.Starner, Phys. Rev. 100, 1390.
55M19 P.Marmier, F.Boehm, Phys. Rev. 97, 103.
55M20 Murray, Boehm, Marmier, DuMond, Phys. Rev. 97, 1007.
55M21 C.E.Mandeville, J.Varma, B.Saraf, Phys. Rev. 98, 94.
55M22 N.Marty, Compt. rend. 240, 291.
55M23 P.Marmier, F.Boehm, private communication quoted in (55A1).
55M24 F.F.Momyer, E.K.Hyde, J. Inorg. Nucl. Chem. 1, 274; NSA 10, 24B, p. 116 (1956).
55M25 Magnusson, Wagner, Engelkemeir, Freedman, ANL-5386.

55N1 R.T.Nichols, E.N.Jensen, Phys. Rev. 100, 1407.
55N2 W.E.Nervik, G.T.Seaborg, Phys. Rev. 97, 1092.
55N3 Nussbaum, Wapstra, Sterk, Kropveld, Physica 21, 77.
55N4 S.J.Nassiff, W.Seelman-Eggebert, Z. Naturf. 10a, 83.
55N5 R.A.Naumann, Phys. Rev. 96, 90.

55O1 Ong, Kramer, Meyer, Fennema, Zijp, Physica 21, 719.
55O2 Ong Ping Hok, P.Kramer, Physica 21, 676.

55P1 Pohm, Waddell, Powers, Jensen, Phys. Rev. 97, 432.
55P2 M.L.Perlman, J.P.Welker, Phys. Rev. 100, 81.
55P3 M.D.Petroff, S.W.Mead, W.O.Doggett, Phys. Rev. 98, 279A.
55P4 A.J.Poe, Phil. Mag. 46, 611.
55P5 C.H.Pruett, R.G.Wilkinson, Phys. Rev. 100, 1237A.
55P6 A.J.Poe, Phil. Mag. 46, 1165.
55P7 Perlman, Asaro, Stephens, Hummel, Pilger, UCRL-2932, p. 59; private communication to NSA 10, 24B, p. 115 (1956).
55P8 I.Perlman, F.Stephens, F.Asaro, Phys. Rev. 98, 262A.
55P9 Prestwood, Smith, Brown, Hoffman, Phys. Rev. 98, 1324.

55R1 Laurence S.Ring, Jr., D.J.Zaffarano, ISC 648.
55R2 H.Roderick, O.Lonsjo, W.E.Meyerhoff, Phys. Rev. 97, 97.
55R3 J.M.Reid, I.F.Wright, Nature 175, 298.
55R4a E.Ricci, J.Pahissa Campa, Nussis, Z. Naturf. 10a, 654.
55R4 L.H.T.Rietjens, H.J.Van den Bold, Physica 21, 701.
55R5 Rollier, Saeland, Morpurgo, Caglieris, Acta Chem. Scandinavica 9, 57; NSA 9, 24B.
55R6 R.C.Rohr, R.D.Birkhoff, Phys. Rev. 98, 1266.
55R7 B.L.Robinson, R.W.Fink, Phys. Rev. 98, 221.
55R8 B.L.Robinson, R.W.Fink, Phys. Rev. 98, 231A.
55R9 N.Ryde, B.Andersson, Proc. Phys. Soc. 68B, 1117.
55R10 R.A.Ricci, G.Trivero, Nuovo Cim. 2, 745.
55R11 R.E.Rowland, Phys. Rev. 99, 757.
55R12 J.O.Rasmussen, H.Slatis, T.O.Passell, Phys. Rev. 99, 42.
55R13 Rasmussen, Stephens, Strominger, Astrom, Phys. Rev. 99, 47.

55S1 A.H.Snell, F.Pleasonton, Phys. Rev. 100, 1396.
55S2 M.J.Sterk, R.H.Nussbaum, A.H.Wapstra, Physica 21, 441.
55S3 Sheline, Wilkinson, Dropesky, Shull, Phys. Rev. 99, 1055.
55S4 A.E.Souch, Proc. Phys. Soc. 68A, 760.
55S5 A.W.Schardt, J.P.Welker, Phys. Rev. 99, 810.
55S6 B.Saraf, J.Varma, C.E.Mandeville, Phys. Rev. 98, 1206A.
55S7 B.Saraf, Phys. Rev. 97, 715.
55S8 P.Staehelin, D.Maeder, M.Pochon, Z. Phys. 140, 498.
55S9 M.J.Sterk, R.H.Nussbaum, H.Cerfontain, Physica 21, 541.
55S10 W.G.Smith, J.M.Hollander, Phys. Rev. 98, 1258.
55S11 F.S.Stephens,Jr., F.Asaro, I.Perlman, Phys. Rev. 100, 1543; J.R.Grover, G.T.Seaborg, ibid.
55S12 D.Strominger, J.O.Rasmussen, Phys. Rev. 100, 844.
55S13 M.H.Studier, J.E.Gindler, C.M.Stevens, Phys. Rev. 97, 88.

55T1 This has been pointed out in TID-5300, p. 47.
55T2 S.Thulin, Arkiv Fysik 9, 137.
55T3 TID-5300, p. 174.
55T4 TID-5300, p. 181.
55T5 G.M.Temmer, N.P.Heydenberg, Phys. Rev. 100, 961A.
55T6 G.M.Temmer, N.P.Heydenberg, Phys. Rev. 98, 1308.
55T7 E.P.Tomlinson, R.A.Naumann, Phys. Rev. 100, 955A.
55T8 H.W.Taylor, R.W.Pringle, Phys. Rev. 99, 1345.

55V1 R.van Lieshout, D.G.Greenberg, C.S.Wu, Phys. Rev. 99, 165.
55V2 J.Varma, C.E.Mandeville, Phys. Rev. 97, 977.

55W1 J.R.Wilkinson, R.K.Sheline, Phys. Rev. 99, 165.
55W2 J.R.Wilkinson, R.K.Sheline, Phys. Rev. 99, 752.
55W3 J.P.Welker, M.L.Perlman, Phys. Rev. 100, 74.
55W4 C.S.Wu, N.Benczer-Koller, private communication (March 1956).
55W5 F.A.White, T.L.Collins, F.M.Rourke, Phys. Rev. 97, 566.
55W6 G.T.Wood, P.S.Jastram, Phys. Rev. 100, 1237A.

55Y1 T. Yuasa, M.E.Nahmiass, M.Vivargent, J. Phys. radium 16, 654.
55Y2 T.Yuasa, J.Laberrigue-Frolow, L.Feuvrais, J. Phys. radium 16, 39, 165.
55Y3 A.I.Yavin, F.H.Schmidt, Phys. Rev. 100, 171.

1956

56A1 E. Arbman, N.Svartholm, Arkiv Fysik 10, 1.
56A2 F.Ajzenberg-Selove, "Classical" Nuclear Physics in the U.S.S.R., Nuvo Cim. Suppl. 4, 2.
56A3 Paul Avignon, Ann. Phys. 1, 10.
56A4 A.H.W.Aten, Jr., Physica 22, 288.
56A5 B.Andersson, Proc. Phys. Soc. 69A, 415.

56B1 Bromley, Gove, Litherland, Paul, Almqvist, BAPS II, 1, 30.
56B2 N.Benczer, C.S.Wu, BAPS II, 1, 41.
56B3 G.E.Boyd, Q.V.Larson, J. Phys. Chem. 60, 707 (O'Kelley, Larson, Boyd, ibid.); NSA 10, 24B.
56B4 W.C.Beggs, B.L.Robinson, R.W.Fink, Phys. Rev. 101, 149.
56B5 A.R.Brosi, B.H.Ketelle, Phys. Rev. 103, 917.
56B6 Bunker, Duffield, Mize, Starner, Phys. Rev. 103, 1417.
56B7 S.K.Bhattacherjee, S.Raman, Nucl. Phys. 1, 486.
56B8 A.Bisi, E.Germagnoli, L.Zappa, Nucl. Phys. 1, 593.
56B9 Ref. 56B7; R.K.Gupta, S.Jha, ibid.
56B10 F.Boehm, E.N.Hatch, BAPS II, 1, 390.
56B11 G.Bertolini, M.Bettoni, E.Lazzarini, Nuovo Cim. 3, 754.
56B12 A.Bisi, S.Terrani, L.Zappa, Nucl. Phys. 1, 425.
56B13 R.Ballini, R.Barloutaud, J. Phys. radium 17, 498.
56B14 A.Bisi, S.Terrani, L.Zappa, Nuovo Cim. 4, 758.
56B15 A.Bisi, E.Germagnoli, L.Zappa, Nuovo Cim. 3, 1007.
56B16 A.Bisi, S.Terrani, L.Zappa, Nuovo Cim. 3, 661.
56B17 F.Boehm, P.Marmier, Phys. Rev. 103, 342.
56B18 G.Backstrom, Arkiv Fysik 10, 387.
56B19 S.K.Bhattacherjee, S.Raman, Nuovo Cim. 3, 1131.
56B20 A.A.Bashilov, Nucl. Phys. 1, 629.
56B21 G.Backstrom, Arkiv Fysik 10, 393.
56B22 S.A.Baranov, K.N.Shliagin, Sov. Phys. JEPT 3, 200.
56B23 Butler, Eastwood, Collins, Jones, Rourke, Schuman, Phys. Rev. 103, 634.
56B24 J.Brunner, J.Halter, H.J.Leisi, Helv. Phys. Acta 29, 235.

56C1 C.S.Cook, F.M.Tomnoves, BAPS II, 1, 31.
56C2 B.Crasemann, D.E.Rehfuss, H.T.Easterday, Phys. Rev. 102, 1344.
56C3 Cork, Brice, Schmid, Helmer, Phys. Rev. 104, 481.
56C4 Cork, Brice, Martin, Schmid, Helmer, Phys. Rev. 101, 1042.
56C5 Cranston, Bunker, Mize, Starner, BAPS II, 1, 389.
56C6 D.R.Connors, W.C.Miller, B.Waldman, Phys. Rev. 102, 1584.
56C7 A.Chetham-Strode, Jr., G.R.Choppin, B.G.Harvey, Phys. Rev. 102, 747.

56D1 Douglas, Gasten, Downey, Mukerji, BAPS II, 1, 21.
56D2 V.S.Dubey, C.E.Mandeville, M.A.Rothman, Phys. Rev. 103, 1430.
56D3 E.der Mateosian, M.Goldhaber, BAPS II, 1, 207.
56D4 M.C.Day, Jr., A.F.Voigt, Phys. Rev. 101, 1784.
56D5 G.T.Danby, A.L.Thompson, J.S.Foster, Proc. Roy. Soc. Canada 50, 27A; NSA 10, 24B (1956).
56D6 V.S.Dubey, C.E.Mandeville, M.A.Rothman, BAPS II, 1, 164.
56D7 P.P.Day, Phys. Rev. 102, 1572.
56D8 Debrunner, Heer, Kundig, Ruetschi, Helv. Phys. Acta 29, 235; Debrunner, Heer, Kundig, Ruetschi, Lindqvist, Helv. Phys. Acta 29, 432.
56D9 F.Demichelis, R.A.Ricci, G.Trivero, Nuovo Cim. 3, 377.
56D10 H.Daniel, Z. Naturf. 11a, 759.
56D11 H.Daniel, R.Nierhaus, Z. Naturf. 11a, 212.
56D12 F.Demichelis, L.A.Radicati, Nuovo Cim. 3, 152.

56E1 G.T.Ewan, M.A.Clark, J.W.Knowles, AECL 343, p. 12 (1956).

56F1 Frauenfelder, Levine, Rossi, Singer, Phys. Rev. 103, 352.
56F2 L.W.Fagg, E.H.Geer, E.A.Wolicki, BAPS II, 1, 165.
56F3 J.Flegenheimer, W.Seelmann-Eggebert, Z. Naturf. 11a, 678.
56F4 Fink, Warren, Edwards, Damon, Phys. Rev. 103, 651.
56F5 Fields, Studier, Diamond, Mech, Inghram, Pyle, Stevens, Fried, Manning, Ghiorso, Thompson, Higgins, Seaborg, Phys. Rev. 102, 180.

56G1 H.J.Gerber, M.G.Munoz, D.Maeder, Phys. Rev. 101, 774; Helv. Phys. Acta 28, 478A (1955).
56G2 W.E.Graves, S.K.Suri, Phys. Rev. 101, 1368.
56G3 M.A.Grace, G.A.Jones, J.O.Newton, Phil. Mag. 1, 363.
56G4 G.Gardener, J.I.Hopkins, Phys. Rev. 101, 999.
56G5 L. Grodzins, H.Motz, Phys. Rev. 102, 761.
56G6 W.E.Graves, A.C.G.Mitchell, Phys. Rev. 101, 701.
56G7 R.K.Gupta, S.Jha, Proc. Phys. Soc. 69A, 70.
56G8 L.Grodzins, H.Kendall, BAPS II, 1, 163, 164.
56G9 T.Gerholm, Arkiv Fysik 11, 55.
56G10 P.R.Gray, Phys. Rev. 101, 1306; G.D.O'Kelley, ibid.

56H1 R.W.Hayward, D.D.Hoppes, Phys. Rev. 104, 183.
56H2 T.H.Handley, S.A.Reynolds, CF-56-65; NSA 10, No. 3243; NSA 10, 24B.
56H3 R.W.Hayward, D.D.Hoppes, Phys. Rev. 101, 93.
56H4 K.Histake, J.T.Jones, J.D.Kurhatov, BAPS II, 1, 271.
56H5 R.W.Hayward, D.D.Hoppes, BAPS II, 1, 42.
56H6 R.W.Hayward, D.D.Hoppes, Phys. Rev. 104, 368.
56H7 B.Hartmann, T.Wiedling, Arkiv Fysik 10, 355.
56H8 E.N.Hatch, F.Boehm, BAPS II, 1, 390.
56H9 Hatch, Marmier, Boehm, DuMond, Phys. Rev. 104, 745.
56H10 J.R.Huizenga, J.Wing, Phys. Rev. 102, 926.
56H11 Herber, Sugihara, Coryell, Bennet, Huizenga, Phys. Rev. 103, 955.
56H12 Huizenga, Engelkemeir, Freedman, Porter, Gindler, BAPS 1, 171.
56H13 R.W.Hoff, J.L.Olsen, L.G.Mann, Phys. Rev. 102, 805.
56H14 J.M.Hollander, W.G.Smith, J.W.Mihelich, Phys. Rev. 102, 740.
56H15 Ref. 56H14; J.O.Newton, private communication to J.O.Rasmussen; ibid.
56H16 D.C.Hoffman, C.I.Browne, J. Inorg. Nucl. Chem. 2, 209; NSA 10, 24B, p. 126 (1956).
56H17 Harvey, Chetham-Strode, Ghiorso, Choppin, Thompson, Phys. Rev. 104, 1315.

56J1 R.G.Johnson, O.E.Johnson, L.M.Langer, Phys. Rev. 102, 1142.
56J2 K.E.Johansson, Arkiv Fysik 10, 247.
56J3 W.C.Jordan, S.B.Burson, J.M.Leblanc, Phys. Rev. 96, 1582.
56J4 R.G.Jung, M.L.Pool, BAPS II, 1, 172.
56J5 N.R.Johnson, R.K.Sheline, R.Wolfgang, Phys. Rev. 102, 831.
56J6 M.C.Joshi, B.V.Thosan, Proc. Indian Acad. Sci. 43A, 255; NSA 10, 24B.
56J7 M.W.Johns, I.R.Williams, D.E.Brodie, Can. J. Phys. 34, 147.
56J8 Johns, McMullen, Williams, Nablo, Can. J. Phys. 34, 69.
56J9 Jones, Schuman, Butler, Cowper, Eastwood, Jackson, Phys. Rev. 102, 203.

56K1 O.C.Kistner, A.Schwarzschild, B.M.Rustad, Phys. Rev. 104, 154.
56K2 Kurbatov, Sathoff, Sakai, Histake, BAPS II, 1, 162.
56K3 Konijn, Van Nooijen, Moster, Endt, Physica 22, 887.
56K4 A.Kjelberg, A.C.Pappas, Nucl. Phys. 1, 322.
56K5 J.J.Kraushaar, E.Brun, W.E.Meyerhof, Phys. Rev. 101, 139.
56K6 S.Katcoff, H.Abrash, Phys. Rev. 103, 966.
56K7 Knight, Mize, Starner, Barnes, Phys. Rev. 102, 1592.
56K8 P. Kleinhaus, Z. Naturf. 11a, 252.
56K9 B.H.Ketelle, H.Thomas, A.R.Brosi, Phys. Rev. 103, 190.
56K10 W.H.Kelly, M.L.Wiedenbeck, Phys. Rev. 102, 1130.
56K11 J.W.Knowles, M.A.Clark, G.T.Ewan, AECL 343, p. 47 (1956).
56K12 S.D.Koicki, A.M.Koicki, Bull. Inst. Nucl. Sci. Boris Kidrich 6, 1.
56K13 P.Kramer, H.C.Hamers, G.Meyer, Physica 22, 208; private communication to NSA 10, 24B, p. 109 (1956).

56L1 W.S.Lyon, N.H.Lazar, Phys. Rev. 101, 1524.
56L2 L.J.Lidofsky, V.K.Fischer, Phys. Rev. 104, 759.
56L3 L.Linder, G.A.Brinkman, Physica 21, 747.
56L4 M.Langevin, Ann. Phys. 1, 57.
56L5 J.Laberrigue-Frolow, Ann. Phys. 1, 152.
56L6 N.H.Lazar, E.Eichler, G.D.O'Kelley, Phys. Rev. 101, 727.
56L7 N.H.Lazar, G.D.O'Kelley, BAPS II, 1, 163.
56L8 M.J.Laubitz, Proc. Phys. Soc. A, 69, 1789.

56L9 J.M.LeBlanc, J.M.Cork, S.B.Burson, Phys. Rev. 104, 1670.
56L10 H.W.Lefevre, E.M.Kinderman, H.H.Van Tuyl, BAPS II, 1, 62; verbal report.

56M1 H.Morinaga, E.Bleuler, BAPS II, 1, 30.
56M2 H.Morinaga, Phys. Rev. 103, 504.
56M3 D.W.Martin, J.M.Cork, S.B.Burson, Phys. Rev. 102, 457.
56M4 D.W.Martin, S.B.Burson, J.M.Cork, BAPS II, 1, 329.
56M5 A.Michalowicz, Compt. rend. 242, 108.
56M6 Martin, Brice, Cork, Burson, Phys. Rev. 101, 182.
56M7 M.Mirnik, A.H.W.Aten, Jr., Physica 22, 14.
56M8 N.Marty, M.Vergnes, Compt. rend. 242, 1438.
56M9 J.W.Mihelich, T.J.Ward, K.P.Jacob, Phys. Rev. 103, 1285.
56M10 J.P.Mize, M.E.Bunker, J.W.Starner, Phys. Rev. 103, 182.
56M11 Th.Mayer-Kuckuk, Z. Naturf. 11a, 627.
56M12 F.F.Momyer, Jr., E.K.Hyde, Phys. Rev. 101, 136.
56M13 J.P.Mize, J.W.Starner, BAPS II, 1, 171; NSA 10, 24B, p. 118 (1956).
56M14 Magnusson, Friedman, Engelkemeir, Fields, Wagner, Phys. Rev. 102, 1097.

56N1 E.Norbeck, Jr., BAPS II, 1, 329.
56N2 Nussbaum, Wapstra, Bruil, Sterk, Nijgh, Grabben, Phys. Rev. 101, 905.
56N3 R.Nordhagen, J.Thomassen, Nucl. Phys. 1, 499.
56N4 D.R.Nethaway, C.M.Michel, W.E.Nervik, Phys. Rev. 103, 147.
56N5 S.V.Nablo, M.W.Johns, BAPS II, 1, 42.
56N6 R.Naumann, BAPS II, 1, 42.

56O1 G.D.O'Kelley, N.H.Lazar, E.Eichler, Phys. Rev. 101, 1059.
56O2 G.D.O'Kelley, N.H.Lazar, E.Eichler, Phys. Rev. 102, 223.
56O3 J.L.Olsen, L.G.Mann, M.Lindner, BAPS II, 1, 41.
56O4 Ref. 56O3; J.M.Hollander, ibid.
56O5 Ong Ping Hok, E.Arbman, Arkiv Fysik 11, 193.
56O6 Ong Ping Hok, J.T.Verschoor, P. Born, Physica 22, 465.

56P1 A.V.Pohm, R.C.Waddell, E.N.Jensen, Phys. Rev. 101, 1315.
56P2 Penning, Maltrud, Hopkins, Schmidt, Phys. Rev. 104, 740.
56P3 A.V.Pohm, R.C.Waddell, E.N.Jensen, Phys. Rev. 101, 1067.
56P4 F.W.Prosser, Jr., M.S.Moore, J.P.Schiffer, BAPS II, 1, 163.
56P5 Perrin, Dick, Foucher, Vartapetian, J. Phys. radium 17, 539 (1956); NSA 10, 24B.
56P6 Michael D.Petroff, BAPS II, 1, 389.
56P7 I.Pullman, P. Axel, Phys. Rev. 102, 1366.
56P8 Porter, Freedman, Novey, Wagner, Jr., Phys. Rev. 103, 921.
56P9 V.R.Potnis, Phys. Rev. 104, 722.
56P10 V.R.Potnis, C.E.Mandeville, J.S.Burlew, Phys. Rev. 101, 753.
56P11 V.R.Potnis, BAPS II, 1, 170.

56R1 T.J.Rock, Proc. Roy. Soc. Canada 50, 28A; NSA 10, 24B (1956).

56S1 A.Schwarzschild, B.M.Rustad, C.S.Wu, BAPS II, 1, 336.
56S2 Sherr, Gerhart, Horie, Hornyak, Phys. Rev. 100, 945.
56S3 T.Stribel, Z. Naturf. 11a, 254.
56S4 B.Saraf, Phys. Rev. 102, 466.
56S5 C.R.Sun, Byron T.Wright, BAPS II, 1, 253.
56S6 A.Schwarzschild, B.M.Rustad, C.S.Wu, Phys. Rev. 103, 1796.
56S7 A.W.Schardt, B.C.Dropesky, BAPS II, 1, 162.
56S8 J.P.Schiffer, M.S.Moore, C.M.Class, Phys. Rev. 104, 1661.
56S9 W.L.Stirling, N.Goldberg, BAPS II, 1, 291.
56S10 R.M.Sinclair, Westinghouse Research Report 60-94511-6-R12; NSA 10, 24B (1956).
56S11 Smith, Gove, Henry, Becker, Phys. Rev. 104, 706.
56S12 E.Schalscha, Nucl. Phys. 1, 481.
56S13 V.S.Shpinel, G.A.Kuznetsova, Sov. Phys. JEPT 3, 216.
56S14 A.Schwarzschild, unpublished data, private communication (March 1957).
56S15 E.Silva, J.Goldemberg, Nuovo Cim. 3, 12.
56S16 D.R.Speck, Phys. Rev. 101, 1725.
56S17 Smith, Hamilton, Robinson, Langer, Phys. Rev. 104, 1020.
56S18 V.W.Shiel, L.D.Wyly, C.H.Braden, BAPS II, 1, 264.
56S19 Schmorak, Stockendal, McDonell, Bergstrom, Gerholm, Nucl. Phys. 2, 193.
56S20 D.Strominger, F.S.Stephens, Jr., J.O.Rasmussen, Phys. Rev. 103, 748.
56S21 K.N.Shliagin, Sov. Phys. JEPT 3, 663.

56T1 N.W.Tanner, Phil. Mag. 1, 47.
56T2 G.M.Temmer, N.P.Heydenberg, Phys. Rev. 100, 961A.
56T3 R.F.Thomas, Jr., W.A.Whitaker, C.L.Peacock, BAPS II, 1, 86.
56T4 W.Turchinetz, R.W.Pringle, Phys. Rev. 103, 1000.
56T5 M.T.Thieme, E.Bleuler, Phys. Rev. 102, 195.
56T6 M.T.Thieme, E.Bleuler, Phys. Rev. 101, 1031.

56V1 J.B.Van der Kooi, H.J.Van den Bold, Physica 22, 681.
56V2 Hamlet Vartapetian, Compt. rend. 243, 1512.
56V3 Vartapetian, Dick, Foucher, Perrin, J.Phys. radium 17, 537.

56W1 J.L.Wolfson, Can. J. Phys. 34, 256.
56W2 R.C.Waddell, E.N.Jensen, Phys. Rev. 102, 816.
56W3 G.W.Warren, R.W.Fink, BAPS II, 1, 171.
56W4 G.K.Wertheim, R.V.Pound, Phys. Rev. 102, 185.

1957

57A1 G.Anderssen, E.Arbman, B.Jung, Arkiv Fysik 11, 297.
57A2 S.Amiel, Phys. Rev. 105, 1412.
57B1 N.Benczer-Koller, A.Schwarzschild, C.S.Wu, BAPS II, 2, 23; private communication (March 1957).
57B2 M.E.Bunker, J.P.Mize, J.W.Starner, Phys. Rev. 105, 227.
57C1 M.A.Clark, J.W.Knowles, BAPS II, 2 (Washington, 1957).
57D1 Dubey, Mandeville, Mukerji, Potnis, BAPS II, 2, 24.
57G1 R.N.Glover, D.E.Watt, Phil. Mag. 2, 49.
57H1 R.L.Heath, Phys. Rev. 105, 1011.
57H2 R.L.Hickok, W.A.McKinley, S.C.Fultz, BAPS II, 2, 24.
57H3 B.C.Haldar, E.O.Wiig, Phys. Rev. 105, 1285.
57J1 R.C.Johnson, R.D.Moffat, BAPS II, 2 (Washington, 1957).
57J2 N.R.Johnson, J.B.Cumming, M.Izawa, BAPS II, 2 (Washington, 1957).
57J3 N.R.Johnson, G.D.O'Kelley, BAPS II, 2, 23.
57J4 S.A.E.Johansson, Phys. Rev. 105, 189.
57K1 Kistner, Schwarzschild, Rustad, Alburger, Phys. Rev. 105, 1339.
57M1 L.G.Mann, R.J.Nagle, H.I.West, BAPS II, 2 (Washington, 1957).
57O1 G.D.O'Kelley, Q.V.Larson, G.E.Boyd, BAPS II, 2, 24.
57P1 J.R.Penning, F.H.Schmidt, Phys. Rev. 105, 647.
57R1 L.P.Roy, J.Roy, J.S.Merritt, Phys. Rev. 105, 1337.

Addenda

55M26 **H.B.Mathur, Phys. Rev. 97, 707 (1955).**

55W7 **J.R.Wilkinson, R.K.Sheline, private communication to TID 5300, p. 55 (1955).**

56G11 **O.O.Gatti, J.Flegenheimer, Z. Naturf. 11a, 679.**

REVIEWS OF MODERN PHYSICS VOLUME 29, NUMBER 4 OCTOBER, 1957

Table of Alpha-Disintegration Energies of the Heavy Elements

FRANK ASARO AND I. PERLMAN

Radiation Laboratory and Department of Chemistry, University of California, Berkeley, California

THIS compilation is a revision of the "Table of Alpha-Disintegration Energies of the Heavy Elements" published in 1954.[1] Included are new alpha emitters and revisions concerning those previously listed. The basis for inclusion in Table II of a previously listed alpha emitter is the availability of additional data which would change the alpha-disintegration energy by more than 1 kev from that listed in Table I.[1] Polonium-211 (0.52 sec) is included but not the 25-sec isomer because it is now certain that the alpha group belonging to the 0.52-sec nuclide represents the transition between ground states.

The only references given are those relevant to the energy determinations. The decay energies are the Q values for the alpha transitions and can be transformed into mass differences by including the atomic mass of He^4.

COLUMN 1

This column indicates the alpha emitter and its product as well as the half-life which is given solely for purposes of further identification. These are the measured half-lives and not the partial alpha-decay half-lives for those cases in which there is more than one mode of decay. Since this table is not a compilation of general decay properties, references are not given for the half-lives cited.

COLUMNS 2 AND 3

In a large fraction of the cases the "highest-energy group" of column 3 is either known to be that of the ground-state transition or is assumed to be so in the absence of information regarding a complex spectrum for the purpose of calculating the disintegration energy of column 2. The Q values, unless otherwise stated under "comments," were calculated by adding to the energy of column 2 the recoil energy, $4E/(A\text{-}4)$, where E is the alpha-particle energy and A is the mass number of the emitter. The Q values were rounded off to values consistent with the precision of the energy measurements.

COLUMN 4

The absence of a notation under "intensity" means that no high-resolution instrument has been used to obtain evidence on complex structure. Otherwise the entry indicates the intensity of the group believed to represent the ground-state transition. The designation "~100" means that a careful search has been made for other groups and either none has been found or that the intensities of lower-energy groups are low.

[1] F. Asaro and I. Perlman, Revs. Modern Phys. 26, 456 (1954).

COLUMN 5

This column refers to the method of energy determination.

ion ch	ionization chamber coupled with some form of pulse-height analyzer.
range air	range determination in air.
range emuls	range of alpha tracks in a photographic emulsion.
spect	magnetic spectrograph.

COLUMN 6

References are given for the energy measurements selected.

COLUMN 7

These letter ratings give the estimated degree of certainty of the isotopic assignments according to the following code:

A Element and mass number certain;
B Element certain and mass number probable;
C Element probable and mass number certain or probable;
D Element certain and mass number not well established.

COLUMN 8

The comments in this column for the most part reinforce the decision on the decay energy.

ins evid	Insufficient evidence to know whether or not the alpha energy measured is that of the ground-state transition.
e-e	No direct evidence, but since the nucleus is of the even-even type it can be assumed that the measured energy is that of the ground-state transition.
α-γ coinc	This designation indicates that coincidences have been observed between alpha particles and gamma rays (or conversion electrons) which show some doubt that the highest-energy alpha group is the ground-state transition. When the evidence is not sufficiently definite to deduce a decay energy based on anything other than the highest-energy alpha group, this is reflected by the values in columns 2 and 3 differing only by the recoil energy. Where the evidence is sufficiently definite to deduce the decay energy, it will be found that columns 2 and 3 differ by more than the recoil energy.

TABLE II.

Reaction	Adopted Q (Mev)	Highest energy group measured (Mev)	Intensity (%)	Method	Energy ref.	Identification	Comments
$Bi^{210} \longrightarrow Tl^{206}$ 2.6×10^{6} y	5.03	4.935 4.97		ion ch ion ch	2 2a	A	
$Bi^{211} \longrightarrow Tl^{207}$ 2.16 m	6.745	6.617 6.620	82.6	spect spect	2b 3	A	
$Po^{197} \longrightarrow Pb^{193}$ ~4 m	6.165	6.040		spect	4	D	ins evid
$Po^{198} \longrightarrow Pb^{194}$ ~6 m	6.057	5.935		spect	4	D	e-e
$Po^{199} \longrightarrow Pb^{195}$ 11 m	5.966	5.846 5.84		spect ion ch	4 4a	B	ins evid
$Po^{200} \longrightarrow Pb^{196}$ ~8 m	5.888	5.770		spect	4	B	e-e
$Po^{201} \longrightarrow Pb^{197}$ 18 m	5.786	5.671 5.70		spect ion ch	4 4b	B	ins evid
$Po^{202} \longrightarrow Pb^{198}$ 51 min	5.689	5.575 5.57 5.61 5.59		spect ion ch ion ch ion ch	4 5 6 7	B	e-e
$Po^{204} \longrightarrow Pb^{200}$ 3.8 h	5.477	5.370 5.37		spect ion ch	4 7	B	e-e
$Po^{211} \longrightarrow Pb^{207}$ 0.52 s	7.58	7.442 7.434	99	spect range air	8 9	A	
$Po^{213} \longrightarrow Pb^{209}$ 4.2×10^{-6} s	8.51	8.35 8.336	~100	spect ion ch	11 13	A	ins evid
$Po^{215} \longrightarrow Pb^{211}$ 1.83×10^{-3} s	7.50	7.360 7.365 7.383	~100	spect range air spect	2b 9 13a	A	
$At^{209} \longrightarrow Bi^{205}$ 5.5 h	5.752	5.642 5.65	~100	spect ion ch	14, 15 17	B	ins evid
$At^{217} \longrightarrow Bi^{213}$ 0.018 s	7.18	7.05 7.02 7.00	~100	spect ion ch ion ch	18 13 19	A	
$Em^{204} \longrightarrow Po^{200}$ 3 m	6.41	6.28		ion ch	5	D	e-e
$Em^{206} \longrightarrow Po^{202}$ 6.2 m	6.37	6.25 6.25		ion ch ion ch	5 6	B	e-e
$Em^{207} \longrightarrow Po^{203}$ 11 m	6.24	6.12 6.09		ion ch ion ch	5 6	B	ins evid
$Em^{208} \longrightarrow Po^{204}$ 23 m	6.261	6.141	~100	spect	20	B	e-e
$Em^{209} \longrightarrow Po^{205}$ 30 m	6.155	6.037		spect	20	B	ins evid
$Em^{210} \longrightarrow Po^{206}$ 2.7 h	6.155	6.037		spect	20	A	
$Em^{212} \longrightarrow Po^{208}$ 23 m	6.384	6.264	~100	spect	20	A	
$Em^{221} \longrightarrow Po^{217}$ 25 m	6.[illegible]	6.0		ion ch	21	A	ins evid
$Em^{219} \longrightarrow Po^{21}$[illegible] 3.92 s	6.940	6.813 6.807	83	spect spect	2b 21b	A	
$Fr^{212} \longrightarrow At^{208}$ 19.3 m	6.534	6.411	37	spect	20	A	ins evid
$Fr^{221} \longrightarrow At^{217}$ 4.8 m	6.449	6.332 6.30	84	spect ion ch	18 13, 21a	A	

Table II.—Continued.

Reaction *	Adopted Q (Mev)	Highest energy group measured (Mev)	Intensity (%)	Method	Energy ref.	Identification	Comments
$Fr^{223} \longrightarrow At^{219}$ 21 m	5.44	5.34		range emuls	22	A	ins evid
$Ra^{223} \longrightarrow Em^{219}$ 11.2 d	5.974	5.867 5.860	9.5	spect spect	2b 22a	A	
$Ra^{222} \longrightarrow Em^{218}$ 38 s	6.671	6.551	96	spect	23	A	
$Ac^{225} \longrightarrow Fr^{221}$ 10.0 d	5.921	5.818 5.80	56	spect ion ch	24 21a, 13	A	
$Th^{226} \longrightarrow Ra^{222}$ 30.9 m	6.444	6.330	79	spect	23	A	
$Th^{227} \longrightarrow Ra^{223}$ 18.6 d	6.144	6.036 6.030	23	spect spect	2b 22a	A	
$Th^{232} \longrightarrow Ra^{228}$ 1.4×10^{10} y	4.077	4.007 4.006		ion ch range emuls	25 26	A	
$Pa^{231} \longrightarrow Ac^{227}$ 3.43×10^{4} y	5.138	5.049 5.046 5.042	8.7	spect spect spect	27 28 29	A	
$U^{230} \longrightarrow Th^{226}$ 20.8 d	5.988	5.884	67.2	spect	23	A	
$U^{233} \longrightarrow Th^{229}$ 1.62×10^{5} y	4.900	4.816 4.823	83.5	spect ion ch	30 13	A	
$U^{234} \longrightarrow Th^{230}$ 2.48×10^{5} y	4.851	4.768 4.768 4.763	72	spect ion ch ion ch	27 25 31	A	
$U^{235} \longrightarrow Th^{231}$ 7.13×10^{8} y	4.67	4.552 4.58	7	spect ion ch	2b 31a	A	
$U^{238} \longrightarrow Th^{234}$ 4.51×10^{9} y	4.267	4.195	77	ion ch	25	A	
$Np^{235} \longrightarrow Pa^{231}$ 410 d	5.25	5.06		ion ch	32	A	α-γ(33)
$Np^{237} \longrightarrow Pa^{233}$ 2.2×10^{6} y	4.950	4.866 4.872	3	spect ion ch	34 35	A	
$Pu^{233} \longrightarrow U^{229}$ 20 m	6.41	6.30		ion ch	36	B	ins evid
$Pu^{236} \longrightarrow U^{232}$ 2.7 y	5.862	5.763	68.9	spect	37	A	
$Pu^{237} \longrightarrow U^{233}$ 44 d	5.75	5.65	21	ion ch	36	A	ins evid
$Pu^{238} \longrightarrow U^{234}$ 89.6 y	5.589	5.495 5.491	72	spect spect	38 27	A	
$Pu^{239} \longrightarrow U^{235}$ 24,360 y	5.238	5.147 5.147 5.150	72.5	spect spect spect	27 39 40	A	isomeric state less than 1 kev 38a, b
$Pu^{240} \longrightarrow U^{236}$ 6580 y	5.246	5.159 5.162	75.5	spect spect	30 40		
$Am^{239} \longrightarrow Np^{235}$ 12 h	5.90	5.78		ion ch	41	A	α-γ(42)
$Am^{241} \longrightarrow Np^{237}$ 461 y	5.626	5.535 5.540	0.42	spect spect	43 27	A	
$Am^{243} \longrightarrow Np^{239}$ 7.9×10^{3} y	5.408	5.339	0.17	spect	44	A	ins evid
$Cm^{238} \longrightarrow Pu^{234}$ 2.4 h	6.63	6.52 6.50		ion ch ion ch	41 45	B	α-e

TABLE II.—*Continued.*

Reaction	Adopted Q (Mev)	Highest energy group measured (Mev)	Intensity (%)	Method	Energy ref.	Identification	Comments
$Cm^{240} \longrightarrow Pu^{236}$	6.38	6.27		ion ch	46	A	e-e
26.8 d		6.25		ion ch	47		
$Cm^{241} \longrightarrow Pu^{237}$	6.20	5.95		ion ch	46	A	α-γ(47a)
35 d							
$Cm^{243} \longrightarrow Pu^{239}$		6.003	1	spect	48	A	
35 y	6.159	5.777	78	spect	83		(49)
$Cm^{245} \longrightarrow Pu^{241}$		5.45	~10	ion ch	50	A	
1×10^4 y	5.62	5.36	~82	ion ch	51		α-γ(48)
		5.4		ion ch	52		
$Cm^{246} \longrightarrow Pu^{242}$	5.46	5.373		ion ch	53	A	e-e
5×10^3 y		5.37		ion ch	50		
		5.39		ion ch	54		
		5.4		ion ch	52		
$Cm^{248} \longrightarrow Pu^{244}$	5.14	5.054		ion ch	53	A	e-e
4.7×10^5 y							
$Bk^{244} \longrightarrow Am^{240}$	6.78	6.67		ion ch	55	B	ins evid
4.35 h							
$Bk^{245} \longrightarrow Am^{241}$	6.48	6.37	33	ion ch	56	A	ins evid
4.95 d		6.35		ion ch	55		
		6.33		ion ch	57		
$Bk^{247} \longrightarrow Am^{243}$	5.85	5.67	~40	ion ch	55	B	α-γ(58)
$\sim10^4$ y							
$Bk^{249} \longrightarrow Am^{245}$	5.53	5.40	~94	ion ch	55	A	α-γ(58)
280 d		5.4		ion ch	59		
		5.4		ion ch	60		
$Cf^{244} \longrightarrow Cm^{240}$	7.29	7.17		ion ch	61	A	e-e
~25 m							
$Cf^{245} \longrightarrow Cm^{241}$	7.23	7.11		ion ch	61	A	
44 m		7.15		ion ch	62		
$Cf^{249} \longrightarrow Cm^{245}$	6.29	6.19	~3	ion ch	63	A	
5×10^2 y		6.19		ion ch	64		
$Cf^{250} \longrightarrow Cm^{246}$	6.122	6.024	83	spect	66	A	
10 y		6.025		ion ch	66a		
		6.033		ion ch	65		
		6.05		ion ch	67		
		6.03		ion ch	59		
$Cf^{252} \longrightarrow Cm^{248}$	6.211	6.112	84.5	spect	66	A	
2.2 y		6.119		ion ch	66a		
		6.117		ion ch	65		
		6.15		ion ch	67		
		6.12		ion ch	59		
$E^{246} \longrightarrow Bk^{242}$	7.4	7.3		ion ch	68	D	ins evid
7.3 m							
$E^{248} \longrightarrow Bk^{244}$	6.98	6.87		ion ch	72	B	ins evid
25 m							
$E^{249} \longrightarrow Bk^{245}$	6.87	6.76		ion ch	73	B	ins evid
2 h							
$E^{251} \longrightarrow Bk^{247}$	6.58	6.48		ion ch	73	B	ins evid
1.5 d							
$E^{252} \longrightarrow Bk^{248}$	6.75	6.64		ion ch	73	B	ins evid
~140 d							
$E^{253} \longrightarrow Bk^{249}$	6.740	6.633	90.2	spect	74	A	
20.03 d		6.636		ion ch	75		
		6.63		ion ch	67		
		6.61		ion ch	76		

TABLE II.—*Continued.*

Reaction	Adopted Q (Mev)	Highest energy group measured (Mev)	Intensity (%)	Method	Energy ref.	Identification	Comments
$E^{254} \longrightarrow Bk^{250}$ ~300 d	6.52	6.42		ion ch	75	A	ins evid
		6.44		ion ch	77		
$Fm^{250} \longrightarrow Cf^{246}$ 30 m	7.55	7.43		ion ch	78	B	e-e
		7.7		ion ch	79		
$Fm^{251} \longrightarrow Cf^{247}$ 7 h	7.00	6.89		ion ch	78	B	ins evid
$Fm^{252} \longrightarrow Cf^{248}$ 23 h	7.16	7.05		ion ch	78	B	e-e
		7.04		ion ch	80		
$Fm^{253} \longrightarrow Cf^{249}$ 4.5 d	7.05	6.94		ion ch	81	B	ins evid
		6.85		ion ch	80		
$Fm^{254} \longrightarrow Cf^{250}$ 3.24 h	7.32	7.20		ion ch	75	A	
		7.22		ion ch	82		
		7.17		ion ch	76		
$Fm^{255} \longrightarrow Cf^{251}$ 21.5 h	7.2	7.08		ion ch	75	B	ins evid
		7.1		ion ch	82		

[1] See reference 1 of text.

[2] H. B. Levy and I. Perlman, Phys. Rev. **94**, 152 (1954).

[2a] M. A. Rollier, Gazz. chim. ital. 84, 658 (1954); Chem. Abstr. 49–12983 f (1955).

[2b] R. C. Pilger, Ph.D. thesis, University of California (University of California Radiation Laboratory Report UCRL-3877) July, 1957.

[3] G. H. Briggs, Revs. Modern Phys. **26**, 1 (1954). This is a compilation of the best values for the energies of natural alpha particles.

[4] S. Rosenblum and H. Tyrén, Compt. rend. **239**, 1205 (1954).

[4a] The authors (see reference 4b) assigned the 5.84 Mev group to Po^{199}. From the work of Rosenblum and Tyrén an assignment to Po^{200} is more reasonable (see reference 4).

[4b] Karraker, Ghiorso, and Templeton, Phys. Rev. **83**, 390 (1951).

[5] A. W. Stoner and E. K. Hyde, J. Inorg. Nuclear Chem. **4**, 77 (1957).

[6] W. E. Burcham, Proc. Phys. Soc. (London) **A67**, 555 (1954).

[7] D. G. Karraker and D. H. Templeton, Phys. Rev. **81**, 510 (1951).

[8] S. Rosenblum, Compt. rend. **193**, 848 (1931). The energy was calculated by Briggs, see reference 3.

[9] W. B. Lewis and B. V. Bowden, Proc. Roy. Soc. (London) **A145**, 235 (1934). (Summarizes the results of various investigators. The values were recalculated according to reference 10.)

[10] M. G. Holloway and M. S. Livingston, Phys. Rev. **54**, 18 (1938). Summarizes work of various investigators.

[11] Stephens, Hummel, Asaro, and Perlman, unpublished data (1954). See also reference 12.

[12] F. S. Stephens, Jr., Ph.D. thesis, University of California (University of California Radiation Laboratory Report UCRL-2970) June 1955.

[13] T. E. Cranshaw and J. A. Harvey, Can. J. Research **26A**, 243 (1948).

[13a] M. Curie and S. Rosenblum, Compt. rend. **194**, 1232 (1932). (This energy was calculated by Briggs, see reference 3.)

[14] R. W. Hoff and F. Asaro, unpublished data (1954). See also University of California Radiation Laboratory Unclassified Report UCRL-3157, p. 67 (September, 1955).

[15] J. P. Hummel and A. W. Stoner, unpublished data (1955); A. W. Stoner, Ph.D. thesis, University of California, University of California Radiation Laboratory Report UCRL-3471 (1956). See also reference 16.

[16] J. P. Hummel, Ph.D. thesis, University of California, University of California Radiation Laboratory Report UCRL-3456 (July 1956).

[17] Barton, Ghiorso, and Perlman, Phys. Rev. **82**, 13 (1951).

[18] Stephens, Hummel, Asaro, and Perlman, unpublished data (1956). See also reference 12.

[19] Hagemann, Katzin, Studier, Ghiorso, and Seaborg, Phys. Rev. **72**, 252 (1947).

[20] Momyer, Asaro, and Hyde, J. Inorg. Nuclear Chem. **1**, 267 (1955).

[21] F. F. Momyer, Jr., and E. K. Hyde, Phys. Rev. **102**, 464 (1956).

[21a] Hagemann, Katzin, Studier, Seaborg, and Ghiorso, Phys. Rev. **79**, 5 (1950).

[21b] Marie Curie and S. Rosenblum, Compt. rend. **196**, 1598 (1933). This energy was recalculated by Briggs (see reference 3) using later values for reference energies.

[22] Jean-Pierre Adloff, Compt. rend. **240**, 1421 (1955).

[22a] Rosenblum, Perey, Valadares, and Guillot; mentioned in Revs. Modern Phys. **25**, 469 (1953) by Hollander, Perlman, and Seaborg.

[23] F. Asaro and I. Perlman, Phys. Rev. **104**, 91 (1956).

[24] Stephens, Hummel, Asaro, and Perlman, unpublished data (1956). See also reference 16.

[25] Harvey, Jackson, Eastwood, and Hanna, Can. J. Phys. **35**, 258 (1957).

[26] Philbert, Génin, and Vigneron, J. phys. radium **15**, 16 (1954).

[27] Goldin, Novikova, and Tretyakov, Conference of the Academy of Sciences, U.S.S.R. on the Peaceful Applications of Atomic Energy, Session of the Division of Physico-Mathematical Sciences 1955, p. 226 (translated by Consultants Bureau, New York, 1955; p. 167).

[28] Hummel, Asaro, and Perlman, Phys. Rev. **98**, 261A (1955).

[29] Rosenblum, Cotton, and Bouissieres, Compt. rend. **229**, 825 (1949).

[30] Goldin, Novikova, and Tretyakov, Phys. Rev. **103**, 1004 (1956).

[31] Clark, Spencer-Palmer, and Woodward. British Declassified Report BR-522, Oct. (1944).

[31a] A. Ghiorso, Phys. Rev. **82**, 979 (1951).

[32] James, Ghiorso, and Orth, Phys. Rev. **85**, 369 (1952).

[33] Hoff, Olsen, and Mann, Phys. Rev. **102**, 805 (1956).

[34] Goldin, Peker, and Novikova, Uspekhi Fiz Nauk **59**, 449 (1956). This is a compilation of heavy-element decay characteristics.

[35] Magnusson, Engelkemeir, Freedman, Porter, and Wagner, Phys. Rev. **100**, 1237A (1955).

[36] Thomas, Vandenbosch, Glass, and Seaborg, Phys. Rev. **106**, 1228 (1957).

[37] Hummel, Asaro, Higgins, and Perlman, unpublished data (1956).

[38] F. Asaro and I. Perlman, Phys. Rev. **94**, 381 (1954).

[38a] F. Asaro and I. Perlman, Phys. Rev. (to be published).

[39] Rosenblum, Valadares, and Goldschmidt, Compt. rend. **230**, 638 (1950).

[40] F. Asaro and I. Perlman. Phys. Rev. **88**, 828 (1952).

[41] G. H. Higgins, Ph.D. thesis, University of California, University of California Radiation Laboratory Unclassified document UCRL-1796, June 1952.

[42] Asaro, Stephens, Gibson, Glass, and Perlman, Phys. Rev. **100**, 1541 (L) (1955).

[43] F. Asaro and I. Perlman, Phys. Rev. **93**, 1423 (1954).

[44] Hummel, Asaro, and Perlman, unpublished data. S[illegible] also reference 16.

[45] Street, Ghiorso, Orth, and Seaborg, unpublished [illegible] October 1948).

[46] R. A. Glass, Ph.D. thesis, University of Calif[illegible] University of California Radiation Laboratory Unclassified Rep[illegible]RL-2560 (April 1954).

[47] G. H. Higgins and K. Street, Jr., Phys. Rev. **86**, [illegible]2 (1952).

[47a] Recoils from the alpha decay of Cm^{241} were found to contain a 145-kev transition with 0.18-sec half-life. Stephens, Asaro, Amiel, and Perlman, Phys. Rev. **107**, 1456 (1957).

[48] Asaro, Thompson, and Perlman, unpublished data (1954).

[49] Hollander, Smith, and Mihelich, Phys. Rev. **102**, 740 (1956). T[illegible] authors deduced the level scheme of Pu^{239} from the decay of Np[illegible].

[50] E. K. Hulet, J. Inorg. Nuclear Chem. (to be published).

[51] Hulet, Thompson, and Ghiorso, Phys. Rev. **95**, 170[illegible] (1954).

[52] Fields, Studier, Diamond, Mech, Inghram, Pyle, Stevens, Fried, Manning, Ghiorso, Thompson, Higgins, and Seaborg, Phys. Rev. 102, 180 (1956).

[53] Butler, Eastwood, Jackson, and Schuman, Phys. Rev. 103, 965 (1956).

[54] Browne, Hoffman, Crane, Balagna, Higgins, Barnes, Hoff, Smith, Mize, and Bunker, J. Inorg. Nuclear Chem. 1, 254 (1955).

[55] A. Chetham-Strode, Jr., Ph.D. thesis, University of California, University of California Radiation Laboratory Unclassified Report UCRL-3322, June 1956.

[56] Magnusson, Friedman, Engelkemeir, Fields, and Wagner, Phys. Rev. 102, 1097 (1956).

[57] Hulet, Thompson, Ghiorso, and Street, Phys. Rev. 84, 366 (1951).

[58] Chetham-Strode, Stephens, and Asaro, unpublished data (1955). See also reference 55.

[59] Diamond, Magnusson, Mech, Stevens, Friedman, Studier, Fields, and Huizenga, Phys. Rev. 94, 1083 (1954).

[60] A. Ghiorso (private communication, 1954).

[61] Chetham-Strode, Choppin, and Harvey, Phys. Rev. 102, 747 (1956).

[62] Thompson, Street, Ghiorso, and Seaborg, Phys. Rev. 80, 790 (1950). The Cf^{245} alpha group was tentatively assigned to Cf^{244} in this paper.

[63] A. Ghiorso and B. G. Harvey, unpublished data (1954). Mentioned in reference 12.

[64] The articles designated in references 65 and 56 give, respectively, the energy of the most abundant alpha group in Cf^{249} decay and the energy separation between this group and the highest energy one.

[65] Magnusson, Studier, Fields, Stevens, Mech, Friedman, Diamond, and Huizenga, Phys. Rev. 96, 1576 (1954).

[66] Asaro, Stephens, Harvey, and Perlman, Phys. Rev. 100, 137 (1955).

[66a] Jones, Schuman, Butler, Cowper, Eastwood, and Jackson, KAPL-1378 (1955).

[67] Ghiorso, Thompson, Higgins, Harvey, and Seaborg, Phys. Rev. 95, 293 (1954).

[68] Ghiorso, Rossi, Harvey, and Thompson, Phys. Rev. 93, 257 (1954). These authors tentatively assigned a 7.35-Mev group to E^{247}. A subsequent investigation (69) indicated the mass number was less than 247. The most likely mass number (70) was thought to be 246. In a later review article (71) the energy was listed as 7.3 rather than 7.35 Mev.

[69] A. Chetham-Strode (private communication, 1956).

[70] A. Ghiorso and S. G. Thompson (private communication, 1956).

[71] Ghiorso, Thompson, Higgins, Harvey, and Seaborg, Phys. Rev. 95, 293 (1954).

[72] A. Chetham-Strode, and L. W. Holm, Phys. Rev. 104, 1314 (1956).

[73] Harvey, Chetham-Strode, Ghiorso, Choppin, and Thompson, Phys. Rev. 104, 1315 (1956).

[74] Hummel, Asaro, Stephens, Thompson, and Perlman, unpublished data (1955).

[75] Jones, Schuman, Butler, Cowper, Eastwood, and Jackson, Phys. Rev. 102, 203 (1956).

[76] Fields, Studier, Mech, Diamond, Friedman, Magnusson, and Huizenga, Phys. Rev. 94, 209 (1954).

[77] Harvey, Thompson, Choppin, and Ghiorso, Phys. Rev. 90, 337 (1955

[78] Amiel, Chetham-Strode, Choppin, Ghiorso, Harvey, Holm, and Thompson, Phys. Rev. 106, 553 (1957).

[79] Atterling, Forsling, Holm, Melander, and Åström, Phys. Rev. 95, 585(L) (1954).

[80] Friedman, Gindler, Barnes, Sjoblom, and Fields, Phys. Rev. 102, 585(L) (1956).

[81] S. Amiel, Phys. Rev. 105, 1412(L) (195_).

[82] Choppin, Thompson, Ghiorso, and Harvey, Phys. Rev. 94, 1080 (1954)

[83] Asaro, Thompson, and Perlman, Phys. Rev. 92, 694 (1953).